



## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

<b>(51) Classification internationale des brevets <sup>6</sup> :</b> C12N 15/13, A61K 39/395, 48/00, C07K 16/18, C12N 15/86, G01N 33/577, 33/68, C12Q 1/68, C12N 15/12	<b>A1</b>	<b>(11) Numéro de publication internationale:</b> WO 97/17445 <b>(43) Date de publication internationale:</b> 15 mai 1997 (15.05.97)
<b>(21) Numéro de la demande internationale:</b> PCT/FR96/01773 <b>(22) Date de dépôt international:</b> 8 novembre 1996 (08.11.96) <b>(30) Données relatives à la priorité:</b> 95/13576 10 novembre 1995 (10.11.95) FR <b>(71) Déposants (pour tous les Etats désignés sauf US):</b> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) [FR/FR]; 3, rue Michel-Ange, F-75016 Paris (FR). INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MÉDICALE (INSERM) [FR/FR]; 101, rue de Tolbiac, F-75013 Paris (FR). <b>(72) Inventeurs; et</b> <b>(75) Inventeurs/Déposants (US seulement):</b> TORA, Lazslo [FR/FR]; 14, rue des Primevères, F-67880 Krautergersheim (FR). LUTZ, Yves [FR/FR]; 12, rue d'Yprès, F-67000 Strasbourg (FR). TROTTIER, Yvon [FR/FR]; 9, rue Kuhn, F-67000 Strasbourg (FR). MANDEL, Jean-Louis [FR/FR]; 9, rue du Barrage, F-67300 Schiltigheim (FR). <b>(74) Mandataires:</b> PEAUCELLE, Chantal etc.; Cabinet Armengaud Aîné, 3, avenue Bugeaud, F-75116 Paris (FR).		<b>(81) Etats désignés:</b> CA, JP, US, brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Publiée</b> <i>Avec rapport de recherche internationale.</i> <i>Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues.</i>
<b>(54) Title:</b> NEURODEGENERATIVE DISEASE TREATMENT AND DIAGNOSTIC MEANS  <b>(54) Titre:</b> MOYENS POUR LE TRAITEMENT ET LE DIAGNOSTIC DE MALADIES NEURODEGENERATIVES  <b>(57) Abstract</b> <p>Means for treating and diagnosing neurodegenerative diseases related to the presence of polyglutamine chains by means of a 1C2 antibody are disclosed.</p> <b>(57) Abrégé</b> <p>La présente invention concerne des moyens pour le traitement et le diagnostic des maladies neurodégénératives associées à la présence de chaînes polyglutaminiques, mettant en œuvre un anticorps 1C2.</p>		

# **UNIQUEMENT A TITRE D'INFORMATION**

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AT	Arménie	GB	Royaume-Uni	MW	Malawi
AT	Autriche	GE	Géorgie	MX	Mexique
AU	Australie	GN	Guinée	NE	Niger
BB	Barbade	GR	Grèce	NL	Pays-Bas
BE	Belgique	HU	Hongrie	NO	Norvège
BF	Burkina Faso	IE	Irlande	NZ	Nouvelle-Zélande
BG	Bulgarie	IT	Italie	PL	Pologne
BJ	Bénin	JP	Japon	PT	Portugal
BR	Brésil	KE	Kenya	RO	Roumanie
BY	Bélarus	KG	Kirghizistan	RU	Fédération de Russie
CA	Canada	KP	République populaire démocratique de Corée	SD	Soudan
CF	République centrafricaine	KR	République de Corée	SE	Suède
CG	Congo	KZ	Kazakhstan	SG	Singapour
CH	Suisse	LI	Liechtenstein	SI	Slovénie
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovaquie
CM	Cameroon	LR	Libéria	SN	Sénégal
CN	Chine	LT	Lituanie	SZ	Swaziland
CS	Tchécoslovaquie	LU	Luxembourg	TD	Tchad
CZ	République tchèque	LV	Lettonie	TG	Togo
DE	Allemagne	MC	Monaco	TJ	Tadjikistan
DK	Danemark	MD	République de Moldova	TT	Trinité-et-Tobago
EE	Estonie	MG	Madagascar	UA	Ukraine
ES	Espagne	ML	Mali	UG	Ouganda
FI	Finlande	MN	Mongolie	US	Etats-Unis d'Amérique
FR	France	MR	Mauritanie	UZ	Ouzbékistan
GA	Gabon			VN	Viet Nam

## MOYENS POUR LE TRAITEMENT ET LE DIAGNOSTIC DE MALADIES NEURODEGENERATIVES

La présente invention concerne des moyens pour le traitement et le diagnostic de maladies neurodégénératives. Elle concerne plus particulièrement l'utilisation d'un anticorps monoclonal capable de reconnaître et d'inactiver les chaînes homopolymères de glutamines dans les protéines spécifiquement associées à ces maladies.

La présence de séquences répétées dans l'ADN est un phénomène connu. Ces séquences peuvent être de différentes natures comme des séquences signal ou enhancer. Il peut également s'agir de séquences codant pour un homopolymère faisant partie d'une structure protéique de plus grande taille.

Dans le cas précis de la maladie de Huntington il s'agit d'une séquence répétée de codons CAG codant pour une chaîne homopolymère de glutamine (polyglutamine). Il a été montré que cette séquence est bien exprimée dans les protéines traduites. L'implication de ces protéines dans le déclenchement ou le développement de la maladie dépend essentiellement du nombre de résidus glutamine enchaînés dans la protéine. Plus celui-ci est important plus la maladie sera sévère et précoce.

On a pour l'instant dénombré au moins cinq maladies neurodégénératives humaines génétiques associées à la présence de ces chaînes de résidus glutamine : l'atrophie musculaire spino-bulbaire associée au chromosome X ou maladie de Kennedy, la maladie de Huntington dominante autosomale, l'ataxie spinocérébelleuse de type 1, l'atrophie dentarorubral-pallidoluysienne et l'ataxie spinocérébelleuse de type 3 ou maladie de Machado-Joseph. Dans les gènes codant pour les protéines responsables de ces maladies, le nombre de triplets CAG répétés est très variable. Par exemple dans le gène responsable de la maladie de Huntington, ce

nombre varie, entre 10 et 35 unités chez les sujets non atteints et de 37-40 jusqu'à 60-120 chez les malades. De plus chez les malades, on observe une instabilité de ce nombre de répétitions d'une génération à l'autre. Une  
5 explication à cette variabilité repose sur les phénomènes de recombinaison et réplication se produisant lors des divisions cellulaires au cours de la gamétogénèse. Ces phénomènes peuvent soit conduire à une augmentation du nombre de répétitions soit, plus rarement, à une  
10 diminution. Dans la plupart des cas le nombre de triplets CAG augmente chez les descendants et l'on observe que cette amplification de taille se fait surtout sur les allèles paternels du gène concerné. Le nombre de recombinaisons subies par l'ADN lors de la spermatogénèse  
15 est en effet plus élevé que celui des recombinaisons survenant lors de l'ovogénèse. Ceci est dû au nombre très élevé de divisions survenant au cours de la spermatogénèse.

Une étude réalisée au sein de plusieurs  
20 familles atteintes de la maladie de Huntington, a permis de comparer sur plusieurs générations quelques paramètres tels que la longueur de ces séquences répétées, l'âge auquel se développe la maladie et la sévérité de celle-ci. Les résultats obtenus font apparaître une corrélation  
25 inverse entre le nombre de triplets CAG (déterminant la longueur de la chaîne polyglutamine) d'une part et l'âge d'apparition et la gravité des symptômes d'autre part. Ceci permet d'expliquer la plus grande précocité et la plus grande sévérité de ces maladies de génération en  
30 génération.

A ce jour il n'existe aucun outil thérapeutique pour le traitement de la maladie de Huntington et d'une manière générale des maladies neurodégénératives associées à une répétition de glutamine.

La présente invention a ainsi pour objet de proposer une méthode de traitement de ces maladies. Elle est fondée sur l'utilisation d'un anticorps capable de se fixer sur les formes pathogènes des protéines responsables des maladies associées à une répétition de glutamine.

Plus précisément, la demanderesse s'est intéressée à la caractérisation d'un système capable de se lier in vitro à des chaînes polyglutamine dont la longueur correspond à celle présente dans les protéines responsables de maladies neurodégénératives. Ceci l'a conduit à rechercher un anticorps monoclonal à même de reconnaître spécifiquement les chaînes polyglutamine contenant un nombre de résidus supérieur à 37 ce qui correspond à la valeur limite inférieure de la longueur de la chaîne polyglutamine dans les protéines pathogènes.

De manière inattendue, la demanderesse a mis en évidence qu'un anticorps monoclonal spécifique, l'anticorps monoclonal 1C2 (mAc1C2) s'avère capable de discriminer les protéines pathogènes des protéines normales en fonction de la longueur de leurs chaînes polyglutamines respectives.

L'anticorps monoclonal 1C2 est déjà connu pour son affinité pour un facteur de transcription se liant aux séquences TATA (TATA-binding protein : TBP). Jusqu'à présent le peptide LEEQQRQQQQQQ, localisé à l'extrémité N-terminale de la chaîne homopolymère de glutamine de la TBP, était considéré comme l'épitope pour lequel l'affinité de cet anticorps était la plus importante (Lescure A et al. *EMBO Journal* 13, 1166-1175 (1995)).

De manière tout à fait surprenante la demanderesse a montré que cet anticorps possédait en fait une très forte affinité pour les séquences polyglutamines même en l'absence du peptide décrit ci-dessus. Cette affinité est, de plus, proportionnelle à la longueur de

la chaîne polyglutamine. Son affinité pour ces dernières est d'autant plus importante que les chaînes sont longues. Pour les chaînes de longueur normale elle est nulle pour un temps d'exposition normal à l'anticorps et  
5 très faible si l'on augmente cette durée. De ce fait l'anticorps 1C2 est capable de reconnaître les longues chaînes polyglutamines des allèles mutés des protéines responsables de la maladie de Huntington et des ataxies spinocérébelleuses 1 et 3 comme épitope pathologique.  
10 Avantageusement, il permet le diagnostic précoce des sujets qui vont développer l'une de ces maladies ainsi que des familles à risques qui expriment des protéines dont la chaîne polyglutamine comprend un nombre de résidus à la limite du pathologique.

15 L'anticorps monoclonal 1C2 reconnaît spécifiquement les formes pathologiques des protéines pathogènes dans la maladie de Huntington et les maladies associées à une répétition de triplets. Il peut être utilisé afin d'inactiver spécifiquement les formes  
20 pathogènes de ces protéines, la liaison de 1C2 pouvant entraîner

- un changement de conformation de la protéine lui faisant perdre ses propriétés pathogènes, ou,

25 - une plus grande sensibilité aux systèmes de dégradation aussi bien intracellulaires qu'extracellulaires.

Un premier objet de la présente invention est par conséquent l'utilisation de l'anticorps 1C2 ou d'un fragment ou d'un dérivé de l'anticorps 1C2 pour la  
30 préparation d'une composition pharmaceutique destinée au traitement préventif ou curatif des maladies neurodégénératives associées à une répétition de glutamine.

35 Les fragments ou dérivés d'anticorps sont par exemple les fragments Fab ou F(ab)'<sub>2</sub>, les régions VH ou

VL d'un anticorps ou encore des anticorps simple chaîne (ScFv) comprenant une région VH liée à une région VL par un bras. Ce type de domaine est particulièrement avantageux puisqu'il peut être dirigé contre toute

5 molécule.

Les anticorps, molécules de la superfamille des immunoglobulines, sont constitués de différentes chaînes (2 lourdes (H) et 2 légères (L)) elles-mêmes composées de différents domaines (domaine variable (V) domaine de

10 jonction (J), etc). Le fragment ou dérivé d'anticorps selon l'invention comprend au moins le site de liaison de l'anticorps aux séquences polyglutamines. Ce fragment peut être soit le domaine variable d'une chaîne légère ( $V_L$ ) ou lourde ( $V_H$ ), éventuellement sous forme de

15 fragment Fab ou  $F(ab')_2$  ou, préférentiellement, sous forme d'anticorps simple chaîne (ScFv). Les anticorps simple chaîne sont constitués d'un peptide correspondant au site de liaison de la région variable de la chaîne légère d'un anticorps relié par un bras peptidique à un

20 peptide correspondant au site de liaison de la région variable de la chaîne lourde d'un anticorps. La construction de séquences d'acides nucléiques codant pour de tels anticorps modifiés selon l'invention a été décrite par exemple dans le brevet US 4 946 778 ou dans

25 les demandes WO 94/02610, WO 94/29446. Ce type de molécule c'est-à-dire comprenant le site de liaison de la région variable de la chaîne légère de l'anticorps 1C2 relié par un bras peptidique au site de liaison de la région variable de la chaîne lourde de l'anticorps 1C2,

30 constitue également un objet de la présente invention.

Pour inactiver lesdites protéines pathogènes l'anticorps peut être administré tel quel dans le système nerveux des patients, par voie stéréotaxique. Dans ce cas l'anticorps sera dirigé contre les molécules

35 pathologiques produites par les cellules malades. La

fixation de l'anticorps entraîne l'inactivation de ces protéines, et entraîne leur dégradation et permet aussi d'éviter leur accumulation à l'intérieur ou l'extérieur des cellules, une des causes possibles de la maladie. Ces anticorps ou des fragments des ces anticorps peuvent également pénétrer à l'intérieur des cellules et ainsi inactiver les protéines qui ne sont pas sécrétées. Ils sont particulièrement avantageux pour le traitement des maladies telles que par exemple la maladie de Huntington, l'ataxie spinocérébelleuse de type 1, 2 ou 3, l'atrophie musculaire spino-bulbaire associée au chromosome X ou maladie de Kennedy, l'atrophie dentarorubral-pallidoluysienne et l'ataxie spinocérébelleuse autosomale dominante.

Un autre mode d'utilisation de l'anticorps consiste à le faire agir directement à l'intérieur de la cellule. Pour ce faire on utilise les méthodes connues de transfert de gènes. Un mode particulier de réalisation de l'invention consiste à faire exprimer dans les cellules du patient un acide nucléique codant pour l'anticorps 1C2 ou pour un fragment ou dérivé l'anticorps 1C2 comme par exemple un fragment ScFv, de cet anticorps.

La séquence d'acides nucléiques codant pour l'anticorps 1C2 ou un fragment ou un dérivé de l'anticorps 1C2 peut être administrée telle quelle, sous forme d'ADN nu selon la technique décrite dans la demande WO 90/11092. Elle peut également être administrée sous forme complexée, par exemple avec du DEAE-dextran (Pagano et al., J. Virol. I (1967) 891), avec des protéines nucléaires (Kaneda et al., Science 243 (1989) 375), avec des lipides (Felgner et al., PNAS 84 (1987) 7413), sous forme de liposomes (Fraley et al., J Biol Chem. 255 (1980) 10431), etc. Préférentiellement, la séquence utilisée dans le cadre de l'invention fait partie d'un vecteur. L'emploi d'un tel vecteur permet en effet



d'améliorer l'administration de l'acide nucléique dans les cellules à traiter, et également d'augmenter sa stabilité dans lesdites cellules, ce qui permet d'obtenir un effet thérapeutique durable. De plus, il est possible  
5 d'introduire plusieurs séquences d'acide nucléique dans un même vecteur, ce qui augmente également l'efficacité du traitement.

Le vecteur utilisé peut être d'origine diverse, dès lors qu'il est capable de transformer les cellules  
10 animales, de préférence les cellules humaines. Dans un mode préféré de mise en oeuvre de l'invention, on utilise un vecteur viral, qui peut être choisis parmi les adénovirus, les rétrovirus, les virus adéno-associés (AAV) ou le virus de l'herpès.

15 A cet égard, la présente invention a également pour objet tout virus recombinant comprenant, inséré dans son génome, un acide nucléique codant pour un fragment ScFv de l'anticorps 1C2. Préférentiellement, les virus utilisés dans le cadre de l'invention sont défectifs,  
20 c'est-à-dire qu'ils sont incapables de se répliquer de façon autonome dans la cellule infectée. Généralement, le génome des virus défectifs utilisés dans le cadre de la présente invention est donc dépourvu au moins des séquences nécessaires à la réplication dudit virus dans  
25 la cellule infectée. Ces régions peuvent être soit éliminées (en tout ou en partie), soit rendues non-fonctionnelles, soit substituées par d'autres séquences et notamment par la séquence codant pour un fragment ScFv de l'anticorps 1C2. Préférentiellement, le virus défectif  
30 conserve néanmoins les séquences de son génome qui sont nécessaires à l'encapsidation des particules virales.

S'agissant plus particulièrement d'adénovirus, différents sérotypes, dont la structure et les propriétés varient quelque peu, ont été caractérisés. Parmi ces  
35 sérotypes, on préfère utiliser dans le cadre de la

présente invention les adénovirus humains de type 2 ou 5 (Ad 2. ou Ad 5) ou les adénovirus d'origine animale (voir demande WO 94/26914). Parmi les adénovirus d'origine animale utilisables dans le cadre de la présente invention on peut citer les adénovirus d'origine canine, bovine, murine, (exemple : Mavl, Beard et al., Virology 75 (1990) 81), ovine, porcine, aviaire ou encore simienne (exemple SAV). De préférence, l'adénovirus d'origine animale est un adénovirus canin, plus préférentiellement un adénovirus CAV2 [(souche manhattan ou A26/61 (ATCC VR-800) par exemple]. De préférence, on utilise dans le cadre de l'invention des adénovirus d'origine humaine ou canine ou mixte.

Préférentiellement, les adénovirus défectifs de l'invention comprennent les ITR, une séquence permettant l'encapsidation et la séquence codant pour un fragment ScFv de l'anticorps 1C2. Encore plus préférentiellement, dans le génome des adénovirus de l'invention, la région E1 au moins est non fonctionnelle. Le gène viral considéré peut être rendu non fonctionnel par toute technique connue de l'homme du métier, et notamment par suppression totale, substitution, délétion partielle, ou addition d'une ou plusieurs bases dans le ou les gènes considérés. De telles modifications peuvent être obtenues in vitro (sur de l'ADN isolé) ou in situ, par exemple, aux moyens des techniques du génie génétique, ou encore par traitement au moyen d'agent mutagènes. D'autres régions peuvent également être modifiées, et notamment la région E3 (WO 95/02697), E2 (WO 94/28938), E4 (WO 94/28152, WO 94/12649, WO 95/02697) et L5 (WO 95/02697). Selon un mode préféré de mise en oeuvre, l'adénovirus selon l'invention comprend une délétion dans les régions E1 et E4. Dans les virus de l'invention, la délétion dans la région E1 s'étend préférentiellement des nucléotides 455 à 3329 sur la séquence de l'adénovirus Ad5.

Les adénovirus recombinants défectifs selon l'invention peuvent être préparés par toute technique connue de l'homme du métier (Levrero et al., Gene 101 (1991) 195, EP 185 573; Graham, EMBO J. 3 (1984) 2917).

- 5 En particulier, ils peuvent être préparés par recombinaison homologue entre un adénovirus et un plasmide portant entre autre la séquence d'ADN codant pour un fragment ScFv de l'anticorps 1C2. La recombinaison homologue se produit après co-transfection
- 10 desdits adénovirus et plasmide dans une lignée cellulaire appropriée. La lignée cellulaire utilisée doit de préférence (i) être transformable par lesdits éléments, et (ii), comporter les séquences capables de compléter la partie du génome de l'adénovirus défectif, de
- 15 préférence sous forme intégrée pour éviter les risques de recombinaison. A titre d'exemple de lignée, on peut mentionner la lignée de rein embryonnaire humain 293 (Graham et al., J. Gen. Virol. 36 (1977) 59) qui contient notamment, intégrée dans son génome, la partie gauche du
- 20 génome d'un adénovirus Ad5 (12%). Des stratégies de construction de vecteurs dérivés des adénovirus ont également été décrites dans les demandes n°FR 93 05954 et FR 93 08596.

- 25 Ensuite, les adénovirus qui se sont multipliés sont récupérés et purifiés selon les techniques classiques de biologie moléculaire, comme illustré dans les exemples.

- Concernant les virus adéno-associés (AAV), il s'agit de virus à ADN de taille relativement réduite, qui
- 30 s'intègrent dans le génome des cellules qu'ils infectent, de manière stable et site-spécifique. Ils sont capables d'infecter un large spectre de cellules, sans induire d'effet sur la croissance, la morphologie ou la différenciation cellulaires. Par ailleurs, ils ne
- 35 semblent pas impliqués dans des pathologies chez l'homme.

Le génome des AAV a été cloné, séquencé et caractérisé. Il comprend environ 4700 bases, et contient à chaque extrémité une région répétée inversée (ITR) de 145 bases environ, servant d'origine de réplication pour le virus.

5 Le reste du génome est divisé en 2 régions essentielles portant les fonctions d'encapsidation : la partie gauche du génome, qui contient le gène rep impliqué dans la réplication virale et l'expression des gènes viraux, la  
10 partie droite du génome, qui contient le gène cap codant pour les protéines de capsid du virus.

L'utilisation de vecteurs dérivés des AAV pour le transfert de gènes in vitro et in vivo a été décrite dans la littérature (voir notamment WO 91/18088; WO 93/09239 ; US 4,797,368. USS.139.941 EP 488 528). Ces  
15 demandes décrivent différentes constructions dérivées des AAV, dans lesquelles les gènes rep et/ou cap sont délétés et remplacés par un gène d'intérêt, et leur utilisation pour transférer in vitro (sur cellules en culture) ou in vivo (directement dans un organisme) ledit gène  
20 d'intérêt. Les AAV recombinants défectifs selon l'invention peuvent être préparés par co-transfection dans une lignée cellulaire infectée par un virus auxiliaire humain (par exemple un adénovirus), d'un plasmide contenant la séquence codant pour un fragment  
25 ScFv de l'anticorps 1C2 bordé de deux régions répétées inversées (ITR) d'AAV. et d'un plasmide portant les gènes d'encapsidation (gènes rep et cap) d'AAV. Les AAV recombinants produits sont ensuite purifiés par des techniques classiques.

30 Concernant les virus de l'herpès et les rétrovirus, la construction de vecteurs recombinants a été largement décrite dans la littérature voir notamment Breakfield et al., New Biologist 3 (1991) 203 : EP 453242, EP178220, Bernstein et al. Genet. Eng.  
35 7(1985)235;Mc Cormick, BioTechnology 3(1985)689, etc.

En particulier, les rétrovirus sont des virus intégratifs, infectant sélectivement les cellules en division. Ils constituent donc des vecteurs d'intérêt pour des applications cancer. Le génome des rétrovirus comprend essentiellement deux LTR, une séquence d'encapsidation et trois régions codantes (gag. pol et env). Dans les vecteurs recombinants dérivés des rétrovirus, les gènes gag, pol et env sont généralement délétés, en tout ou en partie, et remplacés par une séquence d'acide nucléique hétérologue d'intérêt. Ces vecteurs peuvent être réalisés à partir de différents types de rétrovirus tels que notamment le MoMuLV ("murine moloney leukemia virus" : encore désigné MoMLV). le MSV ("murine moloney sarcomavirus"), le HaSV ("harvey sarcoma virus"), le SNV ("spleen necrosis virus") ; le RSV ("rous sarcoma virus") ou encore le virus de Friend.

Pour construire des rétrovirus recombinants selon l'invention comportant un acide nucléique selon l'invention, un plasmide comportant notamment les LTR, la séquence d'encapsidation et ledit acide nucléique est construit, puis utilisé pour transfecter une lignée cellulaire dite d'encapsidation, capable d'apporter en trans les fonctions rétrovirales déficientes dans le plasmide. Généralement, les lignées d'encapsidation sont donc capables d'exprimer les gènes gag. pol et env. De telles lignées d'encapsidation ont été décrites dans l'art antérieur, et notamment la lignée PA317 (US4,861,719), la lignée PsiCRIP (WO90/02806) et la lignée GP+envAm-12 (WO89/07150). Par ailleurs, les rétrovirus recombinants peuvent comporter des modifications au niveau des LTR pour supprimer l'activité transcriptionnelle, ainsi que des séquences d'encapsidation étendues, comportant une partie du gène gag (Bender et al., J. Virol 61 (1987) 1639). Les

rétrovirus recombinants produits sont ensuite purifiés par des techniques classiques.

Pour la mise en oeuvre de la présente invention, il est tout particulièrement avantageux  
5 d'utiliser un adénovirus ou un rétrovirus recombinant défectif. Ces vecteurs possèdent en effet des propriétés particulièrement intéressantes pour le transfert de gènes. L'adénovirus est particulièrement préféré pour le transfert de gènes dans le système nerveux (WO94/08026).

10 Avantageusement, dans les vecteurs de l'invention, la séquence codant pour un fragment ScFv de l'anticorps 1C2 est placée sous le contrôle de signaux permettant son expression dans les cellules nerveuses. Préférentiellement, il s'agit de signaux d'expression  
15 hétérologues, c'est-à-dire de signaux différents de ceux naturellement responsables de l'expression de l'anticorps. Il peut s'agir en particulier de séquences responsables de l'expression d'autres protéines, ou de séquences synthétiques.

20 Notamment, il peut s'agir de séquences promotrices de gènes eucaryotes ou viraux. Par exemple, il peut s'agir de séquences promotrices issues du génome de la cellule que l'on désire infecter. De même, il peut s'agir de séquences promotrices issues du génome d'un  
25 virus, y compris du virus utilisé. A cet égard, on peut citer par exemple les promoteurs E1A, MLP, CMV, LTR-RSV. etc. En outre, ces séquences d'expression peuvent être modifiées par addition de séquences d'activation, de régulation, ou permettant une expression tissu-  
30 spécifique.

La présente invention concerne également toute composition pharmaceutique comprenant soit l'anticorps 1C2, un fragment ou dérivé de cet anticorps soit un ou  
35 plusieurs vecteurs tels que décrits précédemment. Ces compositions pharmaceutiques peuvent être formulées en

vue d'administrations par voie topique, orale, parentérale, intranasale, intraveineuse, intramusculaire, sous-cutanée, intraoculaire, transdermique, intracérébral stéréotaxique, etc. De préférence, les compositions pharmaceutiques de l'invention contiennent un véhicule pharmaceutiquement acceptable pour une formulation injectable, notamment pour une infection directe dans le cerveau du patient. Il peut s'agir en particulier de solutions stériles, isotoniques, ou de compositions sèches, notamment lyophilisées, qui, par addition selon le cas d'eau stérilisée ou de sérum physiologique, permettent la constitution de solutés injectables. L'injection directe dans le cerveau du patient est avantageuse car elle permet de concentrer l'effet thérapeutique au niveau des tissus affectés.

Les compositions selon l'invention sont tout particulièrement utiles pour le traitement des maladies neurodégénératives associées à la présence d'une protéine portant une chaîne homopolymère de glutamine.

Les doses de virus recombinant défectif utilisées pour l'injection peuvent être adaptées en fonction de différents paramètres, et notamment en fonction du vecteur viral, du mode d'administration utilisé, de la pathologie concernée ou encore de la durée du traitement recherchée. D'une manière générale, les adénovirus recombinants selon l'invention sont formulés et administrés sous forme de doses comprises entre  $10^4$  et  $10^{14}$  pfu/ml, et de préférence  $10^6$  à  $10^{10}$  pfu/ml. Le terme pfu ("plaque forming unit") correspond au pouvoir infectieux d'une solution de virus, et est déterminé par infection d'une culture cellulaire appropriée, et mesure, généralement après 48 heures, du nombre de plages de cellules infectées. Les techniques de détermination du titre pfu d'une solution virale sont bien documentées dans la littérature. Concernant les rétrovirus, les

compositions selon l'invention peuvent comporter directement les cellules productrices, en vue de leur implantation.

5 Sur le plan thérapeutique il serait également nécessaire de disposer d'un outil de diagnostic fiable. Un tel outil serait en outre particulièrement avantageux pour le diagnostic des prédispositions familiales à développer ce type de maladies.

10 La présente invention a donc d'autre part pour objet de proposer une méthode de diagnostic de la maladie Huntington fondée sur un test biologique.

15 La demanderesse s'est également intéressée à d'autres applications possibles de l'anticorps 1C2 dans le cadre de l'identification d'agents responsables de maladies neurodégénératives. Bien que les symptômes cliniques manifestés soient souvent très différents la demanderesse a observé que certaines maladies neurodégénératives présentent en revanche de nombreux points communs, quant à leur mode de développement, avec  
20 la maladie de Huntington. Ces ressemblances sont surtout une apparition des symptômes de plus en plus précoces et sévères au cours des générations, notamment mais pas exclusivement par transmission d'un allèle paternel muté. Les agents pathologiques responsables de ces maladies ne  
25 sont pas connus et sont souvent difficilement identifiables. Il est particulièrement intéressant de rechercher s'il existe dans ces maladies un agent pathogène ressemblant dans sa structure à la protéine responsable de la maladie de Huntington. Il est alors  
30 très avantageux d'utiliser l'anticorps monoclonal 1C2 pour détecter chez des sujets atteints de ces maladies la présence de chaînes polyglutamine. Ceci rend possible l'identification des protéines portant ces chaînes qui sont susceptibles d'être les agents pathologiques  
35 recherchés.



1C2 reconnaît ainsi les ataxines-1 à 55 glutamines ou plus, dans le cas des SCA 1 (ataxie spinocérébelleuse 1), et trois protéines (une protéine majeure à 68K et deux protéines mineures à 74K et 87K),  
5 dans le cas des SCA3 (Maladie de Machado-Joseph).

L'anticorps 1C2 permet, en outre, avantageusement de différencier une SCA2 (ataxie spinocérébelleuse de type 2) d'une ADCA II (ataxie cérébelleuse autosomale dominante de type II) en  
10 distinguant les protéines impliquées dans les phénotypes respectifs (protéine de 130K environ pour ADCA II, protéine de 150K environ, pour SCA2, poids moléculaires estimés par migration électrophorétique).

1C2 peut ainsi avantageusement permettre  
15 d'identifier les formes pathogènes des protéines impliquées dans toute maladie neurodégénérative à anticipation prouvée ou suggérée, telle que les ADCA de type I (SCA4, SCA5 par exemple), AD-FSP (paraplégie spastique familiale) ou bien encore dans certaines formes  
20 et dans certains cas de maladies affectives bipolaires (psychoses maniaco-dépressives) ou de schizophrénie.

L'identification des protéines responsables de ces maladies permet d'accéder à l'étape de séquençage. L'invention fournit alors les moyens de construire des  
25 sondes d'ADN appropriées, pour l'identification du gène responsable et la mise en oeuvre de traitements de thérapie génique tels que décrit ci-dessus.

Après avoir caractérisé l'anticorps 1C2 et montré qu'il détecte de manière spécifique sur transferts  
30 Western les protéines pathologiques présentes chez les patients atteints de HD, de SCA1 et de SCA5, la demanderesse a également démontré que des protéines anormales étaient présentes dans des patients de familles SCA2 ou SCA7.

Ceci est en très bonne corrélation avec les observations cliniques d'anticipation chez ces familles. La protéine SCA2 mutante est cytoplasmique avec une masse moléculaire apparente de 150 kDa environ alors que la  
5 protéine SCA7 est nucléaire avec une masse moléculaire d'environ 130 kDa.

La demanderesse a alors utilisé les propriétés surprenantes et avantageuses de l'anticorps 1C2 pour isoler, par criblage d'expression, des gènes impliqués  
10 dans des maladies à extensions polyglutaminiques.

En plus de trois gènes connus, 1C2 a ainsi permis, par criblage de banques d'expression ADNc, de cloner puis de séquencer 6 nouveaux gènes contenant des répétitions CAG et pouvant être impliqués dans des  
15 maladies à chaînes polyglutaminiques (motifs codants les chaînes polyglutaminiques de ces gènes en SEQ ID n°1 à 6 et ADNc de SCA2 en entier en SEQ ID n°7).

Ces six nouveaux gènes ne présentent que de très faibles homologies avec les gènes connus.

20 Un de ces nouveaux gènes (SEQ ID n°7 et 3) porte une mutation chez les patients atteints d'ataxie spinocérébelleuse de type 2 (SCA2), c'est-à-dire liée au chromosome 12q. Ce gène présente une expression ubiquitaire.

25 L'invention, objet de la présente demande, a donc également pour objet six nouveaux gènes susceptibles d'être impliqués dans des maladies neurodégénératives ou psychiatriques à chaînes polyglutaminiques, et, en particulier le gène impliqué dans l'ataxie  
30 spinocérébelleuse de type 2 (gène SCA2).

Les allèles impliqués dans SCA2 ont, dans leur forme normale, de 17 à 29 triplets CAG répétés entre lesquels s'intercalent de 1 à 3 triplet(s) CAA.

35 Dans leur forme mutée, les allèles impliqués dans SCA2 présentent chez les patients étudiés de 37 à 50

triplets CAG répétés, ce nombre n'étant pas limitatif et étant en tout état de cause supérieur à 30 triplets. Ils apparaissent comme particulièrement instables lors des transmissions à la fois paternelles et maternelles. La  
5 séquence de trois d'entre eux présente des chaînes purement CAG.

Le fait qu'une corrélation inverse particulièrement abrupte soit observée entre l'âge où se déclare la maladie et le nombre de répétitions CAG  
10 suggère une plus grande sensibilité à la longueur des chaînes polyglutaminiques pour SCA2 que pour les autres maladies liées à une extension polyglutaminique.

Les expériences précédentes en transferts Western suggéraient que le seuil inférieur de détection  
15 en utilisant l'anticorps 1C2 était d'environ 30 glutamines.

De manière surprenante et inattendue, 1C2 a permis de cloner des ADNc codant pour des chaînes de seulement 12 à 26 glutamines.

20 Cela pourrait être dû à une plus forte sensibilité du clonage d'expression (plus forte abondance locale de protéines cibles et plus faible complexité des autres protéines), et/ou à une différence dans les conditions pour la réaction antigène/anticorps (pas de  
25 dénaturation par le SDS dans le criblage des colonies).

En conséquence l'utilisation de l'anticorps monoclonal 1C2 se généralise au diagnostic précoce de sujets susceptibles de développer toute maladie neurodégénérative liée à l'expression d'une protéine  
30 ayant dans sa structure une longue chaîne polyglutamine, ainsi que des familles à risques qui expriment des protéines dont la chaîne polyglutamine comprend un nombre de résidus à la limite du pathologique. Ce diagnostic peut utiliser directement l'anticorps 1C2 ou se faire par

analyse de l'ADN au niveau de gènes codant pour des polyglutamines, gènes identifiés grâce à l'anticorps 1C2.

Pour procéder à ces différents diagnostics, on utilise un anticorps monoclonal 1C2. Cet anticorps est  
5 mis en contact avec un extrait cellulaire obtenu à partir de cellules du patient exprimant la protéine recherchée. L'anticorps interagit avec cette protéine au niveau de l'épitope représenté par la longue chaîne polyglutamine. Les complexes Anticorps-Antigènes sont ensuite révélés  
10 par tout moyen connu de l'homme du métier (marquage de l'anticorps, utilisation d'un deuxième anticorps fluorescent anti-1C2, ELISA, etc.). L'intensité de l'interaction est déterminée. C'est en fonction de cette dernière que l'on peut établir le diagnostic.

15 1C2 peut également être utilisé pour la localisation subcellulaire des formes pathologiques des protéines impliquées dans la maladie, pour suivre leur accumulation ou l'accumulation de leurs produits dégradation.

20 La méthode objet de la présente invention ouvre la voie à de nouvelles méthodes de traitement des maladies neurodégénératives basées sur une meilleure connaissance des protéines pathologiques qui en sont responsables.

25 L'invention objet de la présente demande, porte non seulement sur l'utilisation des propriétés nouvelles de l'anticorps 1C2 et sur les protéines impliquées dans la pathogénicité des maladies neurodégénératives à chaînes polyglutaminiques, mais aussi sur de nouveaux  
30 gènes également impliqués dans ces maladies.

Les six nouveaux gènes pouvant être impliqués dans des maladies neurodégénératives à chaînes polyglutaminiques selon l'invention, gène SCA2 inclus, sont d'une importance cruciale pour la compréhension des  
35 mécanismes de pathogénicité de ces maladies.

Ils permettent la mise au point directe de sondes d'acides nucléiques, éventuellement marquées de manière à permettre la détection des formes normales ou mutées de ces gènes, capables de s'hybrider avec les  
5 acides nucléiques (ADN ou ARN) impliqués dans ces maladies neurodégénératives à chaînes polyglutaminiques.

De telles sondes nucléiques sont particulièrement utiles pour le suivi des familles à risques, le conseil génétique prénatal et la  
10 discrimination entre les différentes maladies neurodégénératives, certaines d'entre elles pouvant présenter des symptômes proches.

La présente demande a donc également pour objet de telles sondes nucléiques, portant éventuellement une  
15 substitution chimique, sous forme libre ou associée, des compositions pharmaceutiques les renfermant dans un tampon approprié, une méthode in vitro de diagnostic et/ou de conseil génétique mettant en oeuvre lesdites sondes à l'aide d'une technologie telle que PCR, RT-PCR,  
20 et des kits de diagnostic comprenant lesdites sondes.

De telles sondes, selon l'invention, peuvent également servir de vecteurs de substances  
médicamenteuses pour délivrer lesdites substances médicamenteuses au niveau des zones présentant lesdits  
25 gènes, sous leur forme normale ou pathologique.

Les nouveaux gènes selon l'invention permettent également la mise au point directe d'acides nucléiques anti-sens (ADN ou ARN) utiles comme médicaments, dans le  
traitement de maladies neurodégénératives.

30 La présente demande vise donc également de tels acides nucléiques anti-sens, portant éventuellement, le cas échéant, une substitution chimique, sous forme libre ou associée, éventuellement inclus, encapsulé ou adsorbé, des compositions pharmaceutiques les renfermant dans un

tampon approprié, et des kits à usage thérapeutique comprenant lesdits acides nucléiques anti-sens.

La présente demande porte non seulement sur lesdits nouveaux gènes, sondes nucléiques, acides anti-sens selon l'invention mais aussi sur tout acide nucléique présentant une homologie supérieure ou égale à 50% avec ces produits, sur tout fragment de ces produits, et sur toute banque d'acides nucléiques obtenues par criblage d'expression à l'aide de l'anticorps 1C2 de lignées cellulaires, issues de patients ou d'animaux atteints d'une maladie neurodégénérative.

La présente demande vise également un procédé d'identification ou de purification de protéines à chaînes polyglutaminiques utilisant une étape d'immunodétection ou d'immunopurification par l'anticorps 1C2, fragment ou dérivé de cet anticorps, ou pouvant conduire secondairement à identifier le gène correspondant.

La présente demande vise enfin une méthode de diagnostic utilisant l'amplification PCR sur ADN ou RT-PCR sur ARN permettant de détecter des formes mutées dans des gènes codant pour des chaînes polyglutaminiques identifiées ou clonées grâce à l'anticorps 1C2.

La présente invention sera décrite plus en détail à l'aide des exemples qui vont suivre et qui doivent être considérés comme illustratifs et non limitatifs.

Dans ces exemples, il est fait référence aux figures 1 à 8:

- la figure 1 représente le criblage d'expression d'une banque d'ADNc en utilisant l'anticorps 1C2,

- la figure 2 représente la détection par PCR d'allèles étendus dans une famille SCA2,

- la figure 3 représente les structures d'allèles normaux et pathologiques,
- la figure 4 représente la distribution des tailles alléliques au locus SCA2,
- 5       - la figure 5 représente l'instabilité de la répétition SCA2 au cours de transmissions de parents à enfants,
- la figure 6 représente la corrélation entre l'âge de déclenchement de la maladie clinique et le
- 10       nombre de répétitions,
- la figure 7 représente la séquence de l'ADNc SCA2, et
- la figure 8 représente l'analyse par transfert Northern de l'expression du gène SCA2.

15

#### EXEMPLES:

##### Techniques générales de biologie moléculaire

Les méthodes classiquement utilisées en

20 biologie moléculaire telle que la technique de Western Blot, le marquage d'anticorps, etc... sont bien connues de l'homme de métier et sont abondamment décrites dans la littérature (Maniatis T. et al., "Molecular Cloning, a Laboratory Manual", Cold Spring Harbor Laboratory, Cold

25 Spring Harbor, N.Y., 1982; Ausubel F.M.. et al, (eds), "Current Protocols in Molecular Biology", John Wiley & Sons, New York, 1987).

30 Exemple 1: Mise en évidence de l'affinité spécifique de l'anticorps monoclonal 1C2 pour les chaînes polyglutamines.

Afin de mettre en évidence sa capacité à reconnaître des protéines possédant une chaîne homopolymère de glutamine, mAc1C2 a tout d'abord analysé

35 en "Western Blot" sur des extraits de lignées cellulaires

lymphoblastoïdes (ci-après LCL) provenant d'individus normaux et d'individus atteints de la maladie de Huntington présentant des longueurs variées de chaînes polyglutamines dans les HDP (Huntington Disease Protein).

5 Ces différentes protéines sont tout d'abord mises au contact d'un anticorps monoclonal anti-HDP. On observe lors de l'analyse en Western Blot que l'on discrimine très facilement les protéines normales de celles ayant une chaîne polyglutamine allongée, grâce à  
10 leur différence de poids moléculaire. Par contre, lorsque le même lot est analysé avec l'anticorps monoclonal 1C2, seules les protéines HDP à longue chaîne polyglutamine sont détectées, les protéines normales ne donnant pas de réponse. La réponse obtenue avec mAc1C2 est spécifique  
15 des protéines pathologiques.

Il est également très intéressant de constater que l'intensité du signal dépend de la longueur de la chaîne polyglutamine. Cette intensité est très forte pour les chaînes de plus de 50 unités et minimale pour des  
20 chaînes de 39-40 unités.

Exemple 2: Mise en évidence d'une corrélation entre la longueur de la chaîne polyglutamine et l'augmentation de l'affinité de 1C2 pour celle-ci

Cette relation entre intensité du signal et  
25 longueur de chaîne a ensuite été examinée à l'aide d'échantillons ordonnés en ordre de longueurs de chaînes décroissantes en partant du plus grand allèle. On observe que la valeur de l'affinité de l'anticorps monoclonal 1C2 pour les formes mutantes de HDP dépend clairement de la  
30 longueur de la chaîne de polyglutamine. En effet, l'intensité du signal est plus forte pour les protéines ayant une chaîne de 60 à 85 résidus que pour les protéines ayant une chaîne de 39 ou 40 résidus, cette dernière étant elle-même plus forte que celle observée  
35 pour des chaînes plus courtes.



On observe également qu'une durée d'exposition à l'anticorps prolongée permet la détection des protéines ayant une chaîne dont la longueur se situe à la limite supérieure des protéines normales (> 28).

5                    Une comparaison semi-quantitative de l'intensité du signal détecté avec des HDP comprenant 36, 60 et 85 résidus glutamine a été effectuée à partir d'une série de dilutions d'extraits de LCL. L'intensité du signal HDP observée est 2 à 4 fois plus importante avec  
10 des chaînes de 85 résidus qu'avec des chaînes de 60 résidus et de 10 à 20 fois plus importante avec des chaînes de 60 résidus qu'avec des chaînes de 39 résidus.

15                    Exemple 3: Mise évidence que l'épitope reconnu par mAc1C2 est uniquement la chaîne polyglutamine

Lorsqu'on procède au séquençage de la protéine HDP on remarque que le peptide LEEQQRQQQQQQ de TBP qui est reconnu par l'anticorps n'est pas présent dans les séquences qui entourent la chaîne de résidus glutamine  
20 dans HDP. On en déduit que l'épitope de HDP qui est reconnu par l'anticorps monoclonal 1C2 est bien uniquement la chaîne polyglutamine et que l'intensité du signal est uniquement dépendante de la longueur de celle-ci.

25                    Une expérience de vérification a été faite avec différents allèles de TBP qui possèdent également des chaînes polyglutamine dont la longueur varie de 29 à 42 résidus. L'analyse en Western Blot, après exposition à 1C2, a permis de discriminer les allèles en fonction de  
30 leur taille. Là aussi l'intensité du signal est plus importante avec les chaînes de grandes tailles et les chaînes comprenant entre 27 et 30 résidus ne sont pas détectées.

35                    L'anticorps 1C2 est donc capable de reconnaître spécifiquement une séquence polyglutamine. L'intensité du

signal obtenu dépend de la longueur de ladite séquence plus celle-ci est longue plus l'intensité est forte.

5     Exemple 4: Détection d'une épitope pathologique  
dans les ataxies spinocérébelleuses SCA1 et SCA3.

10     Pour mettre en évidence la capacité de mAclC2 à détecter sélectivement d'autres protéines pathogènes comprenant une longue chaîne polyglutamine, nous avons analysé en Western Blot des extraits de LCL provenant de patients atteints de SCA1 et SCA3.

15     Lors de l'expérience de liaison avec mAclC2, une protéine de 100 kD a été spécifiquement détectée dans les extraits provenant de patients atteints de SCA1 alors qu'elle était absente des extraits provenant de patients atteints de SCA3. Cette protéine correspond à l'ataxine 1, la protéine responsable de SCA1. Inversement, dans les extraits de LCL provenant de patients atteints de SCA3 on a détecté au moins 4 protéines (une bande majeure correspondant à une protéine de 68 kD et trois bandes mineures correspondant à des protéines de 64, 74 et 87 kD) qui sont absentes des extraits provenant de patients atteints de SCA1. Un contrôle effectué sur 9 LCL provenant de sujets sains ne montre aucune de ces bandes. Dans tous les cas, aussi bien chez les sujets sains que chez les sujets atteints, on retrouve une bande correspondant à la TBP et une autre correspondant à une protéine d'environ 230 kD. On peut en conclure que mAclC2 est spécifiquement des protéines responsables de SCA1 et SCA3 et peut donc être utilisé dans le diagnostic de ces deux maladies.

Exemple 5: Mise en évidence de la présence de protéine contenant une longue chaîne polyglutamine dans d'autres maladies neurodégénératives grâce à mAclC2.

Les caractéristiques phénotypiques communes à toutes les maladies considérées se retrouvent dans d'autres maladies neurodégénératives comme l'ataxie cérébelleuse autosomale dominante (ADCA) et la paraplégie spasmodique familiale (FSP) pour lesquelles la/les protéine(s) responsables ainsi que le(s) gène(s) correspondant n'ont pas encore été mis en évidence.

On a testé en aveugle des extraits de LCL provenant de sujets atteints de ces deux maladies, et de SCA1, SCA3 et la maladie de Huntington.

Sur les 9 LCL testées, 4 montrent une bande correspondant à une protéine spécifique de 130 kD ou de 150 kD, 3 ne montrent aucun signal spécifique et les deux autres présentent des bandes correspondant aux protéines mutées responsables de SCA1 et appartiennent aux témoins.

Les résultats obtenus ont été comparés aux données des dossiers médicaux des sujets de l'expérience. Les colonnes portant la bande pour une protéine de 130 kD correspondent à des échantillons provenant de patients atteints d'ADCA de type II entraînant une dégénération rétinienne. Des expériences de cartographie chromosomique ont permis de localiser le gène correspondant à cette protéine de 130 kD sur le chromosome 3p, ce locus correspond au locus présumé responsable de la maladie. On a également retrouvé cette même protéine dans d'autres patients atteints d'ADCA de type II appartenant à d'autres familles.

La protéine de 150 kD a été recherchée chez d'autres patients atteints d'ADCA portant une mutation dans le gène SCA2 ainsi que chez des personnes de la même famille n'ayant pas développé la maladie. La protéine est toujours présente mais chez les sujets sains la longueur de la chaîne polyglutamine est plus petite et donc non détectée dans les conditions expérimentales telles que décrites ci-dessus.

On peut donc en déduire que cette protéine de 150 kD est bien responsable d'une maladie neurodégénérative associée à l'allongement d'une chaîne polyglutamine dans une protéine normale qui présente les mêmes caractéristiques de transmission que la maladie de Huntington.

Exemple 6: localisation intracellulaire des protéines pathologiques.

L'anticorps mAclC2 a également été utilisé pour déterminer la localisation intracellulaire de la protéine responsable de SCA3 ainsi que celle des protéines nouvellement identifiées et qui sont liées à l'ADCA de type II et à SCA2.

L'analyse des fractions cellulaires est effectuée selon la technique du Western Blot. On utilise des fractions enrichies provenant des différents compartiments cellulaires : le compartiment cytoplasmique et le compartiment nucléoplasmique. L'hybridation est réalisée avec mAclC2 marqué.

Les ataxines 2 et 3 ainsi que la HDP mutante ont été localisées dans la fraction cytoplasmique. La protéine de 130 kD liée à l'ADCA de type II a quant à elle été localisée dans la fraction nucléoplasmique. Des tests de contrôle effectué avec la TBP dont la localisation cellulaire est connue ont permis de valider ces résultats.

Exemple 7: clonage du gène impliqué dans l'ataxie cérébelleuse de type 2 (SCA2).

Méthode

**Banques d'expression ADNc**

Des ARN poly A+ SCA2 et SCA7 ont été préparés à partir de lignées cellulaires lymphoblastoïdiques (LCL)

de patients SCA2 et SCA7. La transcription inverse a été réalisée en utilisant des oligonucléotides hexamériques aléatoires. Les ADNc ont été ligaturés à des adaptateurs EcoRI et clonés dans des bras de vecteur EcoRI I-SCREEN-1  
5 (Novagen®) et insérés en suivant le protocole du fabricant.

**Criblage des banques d'expression en utilisant l'anticorps 1C2.**

10 Les phages ont été incubés pendant 15 minutes pour infection avec les bactéries BL21 dans LB contenant du maltose 0,2 mM, MgSO<sub>4</sub> 10 mM et du chloramphénicol à 40 mg/ml.\* Environ  $8 \cdot 10^5$  pfu de chaque banque ont été déposées sur un milieu NZY. Lorsque les plaques furent  
15 visibles, une membrane en nylon imbibée de IPTG 10mM a été placée sur les boîtes et l'induction de l'expression a été réalisée pendant 3 heures 30 minutes. Les membranes ont été ensuite lavées deux fois dans du PBS 1x, Tween 0,05% pendant 5 et 30 minutes respectivement.

20 Les membranes ont été bloquées dans du lait écrémé à 5% puis incubées avec l'anticorps monoclonal 1C2. L'anticorps secondaire (immunoglobulines de chèvre anti-souris) a été couplé à une peroxydase pour permettre une révélation avec le kit ECL (Amersham®). Les temps  
25 d'exposition sont couramment de 20 minutes.

Les plaques positives ont été éluées à 4°C dans un milieu SM et les phages ont été transformés dans les bactéries BL21 comme ci-dessus décrit. Un criblage secondaire et, si nécessaire, tertiaire a été réalisé  
30 comme ci-dessus. Les phages positifs isolés ont été excisés en suivant le protocole du fabricant (Novagen®) et les plasmides obtenus ont été transformés dans des bactéries HB101.

35 **Analyses PCR et séquençage direct**

Des RT-PCR ont été réalisées en utilisant les amorces suivantes:

DAN1: CGTGCGAGCCGGTGTATGGG (UH13); GGCGACGCTAGAAGGCCGCT (UH1  
5 DAN15: CCACCATGCCACACCTCC; CCGCGCCGCCCAAGCTGTTG;  
DAN26: AATGACGTGCTGCACCACTG; CCAGGCATCTGGATGGGAGG;  
AAD10: CCTCGGACCTGATTCAAGGC; GCTGCTGGGAGGCATAAGGC;  
AAD14: AAGTGCCCCTGTCCATCCTCT; GGAGAGGAGTGCAACAGACC;  
10 AAD20: CGGTCGCGGCAATCCTAG; GAGGTTCCGGCTCGGACT.

Les amorces réalisées pour DAN1, DAN15, AAD14 et AAD20 permettent l'amplification de l'ADN génomique.

15 Les produits ont été analysés sur un gel d'agarose à 2% et transférés sur une membrane de nylon pour hybridation avec une sonde oligonucléotidique (CAG)16. Pour l'analyse PCR des allèles SCA2, 100 ng  
20 d'ADN génomique ont été amplifiés dans 20 µl de Tris-HCl 10mM (pH 8,3), KCl 50 mM, MgCl<sub>2</sub> 1,5 mM, glycérol 15%, 250 mM de chaque dNTP, et 10 pmoles d'amorces UH10 et UH13. Après un démarrage à chaud de 5 minutes à 96°C, 0,5 U de Taq polymérase a été ajoutée, 35 cycles (30 secondes à 94°C, 30 secondes à 65°C, 30 secondes à 72°C) et une  
25 élongation finale de 10 minutes à 72°C ont été réalisés. Les produits ont été déposés sur un gel dénaturant à 6% de polyacrylamide et 7 M d'urée. Ils sont transférés et hybridés avec une sonde (CAG)7. Pour le séquençage direct des allèles SCA2, 40 cycles ont été réalisés (les allèles  
30 "étendus" ont été mieux amplifiés en l'absence de KCl et de glycérol) et les fragments excisés ont été séquencés sur un séquenceur automatique Applied Biosystems® (ABI) avec des didésoxynucléotides fluorescents.

35

#### Analyses des transferts Northern et Zoo

Un fragment d'ADNc EcoRI de 2,5 kb correspondant à l'extrémité 3' du clone DAN1 a servi de sonde pour les analyses de transferts Northern et Zoo. Les transferts Northern de MTN humains et MTN II issus de cerveaux humains ont été obtenus par Clontech®. L'hybridation des sondes a été réalisée en utilisant la solution d'hybridation ExpressHyb® (Clontech®). Les transferts ont été lavés dans du SSC 0,1x, SDS 0,1% à 55°C. L'hybridation des transferts Zoo a été réalisée dans du formamide à 30%. Le lavage a été réalisé dans du SSC 0,5 x, SDS 0,1% à 51°C.

### . Résultats

#### 15 Clonage d'expression d'ADNc contenant des polyglutamines

L'utilisation selon l'invention de l'anticorps monoclonal 1C2 avait permis précédemment de détecter dans des lignées lymphoblastiques de patients présentant les formes SCA2 et SCA7 de ADCA (ataxie cérébelleuse autosomale dominante) des protéines pathologiques devant contenir une longue chaîne polyglutaminique.

Pour essayer de cloner les ADNc correspondants, deux banques d'ADNc lymphoblastiques ont été construites dans un vecteur d'expression de bactériophages I (I-Screen-1, Novagen®), en utilisant des lignées cellulaires de patients SCA2 (AAD) ou SCA7 (DAN). Environ  $8 \cdot 10^5$  clones ont été déposés sur plaques à partir de chaque banque et ont été criblés avec l'anticorps 1C2.

21 clones positifs ont été obtenus après trois itérations de criblage.

En figure 1, est représenté le criblage d'expression d'une banque d'ADNc en utilisant l'anticorps 1C2: la partie supérieure gauche rectangulaire montre la détection du clone DAN1 à l'étape primaire de criblage (environ 20 000 pfu par boîte), avec un arrière-plan très

clair. Les autres parties rectangulaires correspondent aux criblages secondaires de ce clone (environ 100 pfu par boîte).

- 5 Tous, excepté deux, contiennent des répétitions CAG telles que déterminées par hybridation avec une sonde oligonucléotidique (CAG)<sub>10</sub>.

Les 19 clones positifs à CAG correspondent à 8 transcrits différents et sont présentés dans le tableau suivant.



Tableau 1: clones obtenus par criblage avec 1C2

clones	taille	polygln	motifs codants polygln	analogues sur banques de données
nouveaux gènes				
AAD10	0.7 kb	11 et 8	(CAG)6 CAA (CAG)4 (NNN)8 (CAG)2 CAA (CAG)3 (NNN)97 (CAG)5 (CAA)2 CAG CAC CAA CAG CAA	X85325 (i)
AAD14	1.9 kb	12 et 14	(CAG)7 CAA (CAG)4 (NNN)25 (CAG)14	M62043; F11363; STS G16005 6 ESTs aucune M85636; R72355; H46162; H43176
DAN1	4 kb	22	(CAG)8 CAA (CAG)4 CAA (CAG)8	
DAN15	4 kb	18	(CAG)7 CAA (CAG)7 CAA (CAG) CAA	
DAN26	1.8 kb	12	(CAG)6 CAA (CAG)5	
AAD20	1.9 kb	(14) (ii)	(CAG)14	M65150 (glutaminase de rat)
gènes connus				
AAD5	Huntington	20	(CAG)18 CAA CAG	
AAD38	hSNFalpha	23 to 26	CAG (CAA)2 (CAG)3 CAA (CAG)13 CAA (CAG)2 CCA (CAG)3 (NNN)6 CAG (CAA)2 (CAG)2	
DAN20	actine beta	aucune		

- (i) l'extrémité 3' de AAD10 chevauche l'extrémité 5' de X85325 ce qui a été décrit comme contenant une chaîne glutaminique non polymorphique interrompue. Cette chaîne apparaît comme étant dans le cadre de lecture avec les régions codantes polyglutaminiques de AAD10.
- (ii) le plus grand cadre de lecture ouvert de AAD20 prédit que la répétition code pour une polyleucine.
- (iii) excepté pour G16005, l'homologie ne chevauche pas les répétitions CAG.

Les 19 clones positifs incluent un ADNc d'huntingtine normale (20 gln) et un ADNc pour le facteur de transcription hSNF- $\alpha^{32}$ , qui contient aussi une répétition polyglutaminique (26 gln). La taille de la dernière protéine (174 kDa) l'exclut comme candidat pour les gènes de SCA2 ou SCA7.

Les six autres transcrits correspondent à des gènes contenant CAG nouveaux.

Aucun des clones ne présente cependant un nombre de répétitions pathologique attendu ( > 35 gln), les plus longues chaînes consistant en 14 CAG ininterrompus (AAD14 et AAD20). Dans tous les gènes exceptés AAD20, le cadre de lecture ouvert permet de prédire de manière non ambiguë une chaîne polyglutaminique. En tenant compte des interruptions par les codons CAA, les plus longues chaînes polyglutaminiques prédites ont été trouvées dans DAN1 (22 gln) et DAN15 (18 gln). Ce résultat est inattendu par rapport aux résultats obtenus au préalable par transferts Western pour lesquels 1C2 ne détecte que les chaînes d'une longueur supérieure à 30 gln dans des extraits cellulaires entiers.

#### Identification d'un clone SCA2.

Bien que les répétitions obtenues n'apparaissent très étendues, il restait possible qu'un des clones représente un allèle normal au locus SCA2 ou SCA7, ou bien la rétractation d'un allèle étendu du fait de l'instabilité des longues répétitions dans les bactéries.

Des amorces adjacentes aux répétitions (voir méthode) ont donc été construites et testées par RT-PCR ou PCR sur, respectivement, de l'ARNm ou de l'ADN génomique obtenus de patients SCA2 et SCA7.

Une paire d'amorces (dérivées de DAN1) détecte les fragments RT-PCR étendus des patients SCA2. Cela a été confirmé en utilisant les mêmes sondes dans toutes les familles SCA2 testées au niveau de l'ADN génomique.

5 En figure 2 est représentée la détection par PCR d'allèles étendus dans une famille SCA2: l'analyse PCR a été réalisée en utilisant les sondes UH10 et UH13; les tailles alléliques (nombre de répétitions) sont 22/37, 23/38, 22/43, 22/43, 22/23 et 22/22 pour les  
10 individus 1 à 6 respectivement; le père dans la génération II a transmis des allèles étendus par 5 répétitions à ces deux filles affectées; ceci est corrélé avec une forte anticipation (l'âge de déclenchement de la maladie est indiqué sur le pedigree); on note une  
15 hétérogénéité apparente des allèles mutants; la plus forte bande a été utilisée pour la détermination des tailles de répétitions.

De plus, ces amorces amplifient le fragment correspondant dans quatre YAC de la région candidate SCA2  
20 de 12q23-24.1 (CEPH YAC 674f2, 722h7, 774a3 et 910g1) qui contiennent également le microsatellite D12S1340 (AFM291xe9).

#### Allèles normaux et pathologiques

25 La répétition originelle dans le clone DAN1 est interrompue par deux triplets CAA. Afin de vérifier si les interruptions sont trouvées dans les allèles normaux de manière générale, les produits PCR correspondants à 17 allèles normaux indépendants ont été séquencés.

30 Tous les allèles analysés contiennent de 1 à 3 CAA dispersés, dans les plus communs des cas (9 sur 17) deux CAA sont observés avec une structure (CAG)8 CAA (CAG)4 CAA (CAG)8, comme pour DAN1.

35 En figure 3, est représentée la structure d'allèles normaux et pathologiques: le séquençage direct

des produits PCR montre des codons CAA intercalaires (cercles pleins) parmi des répétitions CAG (cercles vides); deux allèles pathologiques à 40 et 41 répétitions et un allèle à 34 répétitions d'un porteur cliniquement normal d'un haplotype SCA2 ont été amplifiés dans différentes conditions et ne montrent pas de CAA intercalaires.

Par analyse PCR de 110 allèles, un nombre normal de 17 à 29 répétitions a été observé. 22 répétitions ont été observées pour 75% des allèles alors que le nombre lié à la pathogénicité était de 37 à 50 répétitions (n=31).

En figure 4, est représentée la distribution des tailles alléliques au locus SCA2: les analyses PCR d'individus français normaux et SCA2 ont été réalisées comme décrit ci-dessus; les 110 allèles normaux sont des allèles indépendants alors que les 31 allèles pathologiques (barres pleines) dérivent de 8 familles.

Deux allèles mutants indépendants ont été séquencés et se présentent comme constitués de pures chaînes CAG (voir figure 3), tout comme s'est présenté un allèle à 34 CAG d'un individu âgé de 32 ans cliniquement normal issu d'une famille SCA2. Cet individu est porteur de l'haplotype pathologique.

#### Instabilité et âge de déclenchement de la maladie

Dans les 16 cas de transmissions de parents à enfants que nous avons pu étudier, 13 ont mis en évidence une instabilité. Un seul cas a mis en évidence une diminution (de 3 unités), et de manière plus frappante, 5 cas ont mis en évidence une augmentation de 5 à 10 répétitions. Voir, à ce sujet, la figure 2 et voir également la figure 5 qui représente l'instabilité de la répétition SCA2 lors d'une transmission de parents à

enfants: la différence entre le nombre d'unités de répétitions dans les allèles étendus de parents et enfants a été déterminée pour neuf transmissions paternelles et sept transmissions maternelles chez 8 familles.

Ces grands sauts ont été observés pour les transmissions à la fois paternelles et maternelles, contrairement à ce qui a été observé dans les autres maladies liées aux chaînes polyglutaminiques.

Une forte corrélation inverse entre l'âge de déclenchement de la maladie et la longueur des répétitions a été observée pour 26 patients ( $r = -0,86$  avec une régression quadratique,  $p = 0,0001$ ).

En figure 6, est représentée la corrélation entre l'âge de déclenchement de la maladie clinique et le nombre de répétition: les données proviennent de 26 patients (âge moyen de déclenchement de la maladie = 34 ans, gamme allant de 13 à 60 ans); le coefficient de corrélation est calculé pour une régression quadratique ( $r = -0,86$ ;  $p < 0,0001$ ).

L'effet de répétitions additionnelles est frappant: pour quatre patients présentant 37 répétitions, l'âge de déclenchement de la maladie s'est situé entre 45 et 60 ans alors que pour les trois patients présentant de 46 à 50 répétitions, l'âge de déclenchement de la maladie s'est situé entre 13 et 18 ans.

#### Le gène SCA2 et son expression

Le clone DAN1 (4,0kb) a été entièrement séquencé: en figure 7, est représentée la SEQ ID n°7.

La figure 7 représente en effet la séquence de l'ADNc SCA2. La séquence nucléotidique de la position 1 à 3986 provient du clone DAN1. Les dernières 177 paires de bases (en italique) proviennent de EST (H92640, N90240 et Z13574 de dbEST) qui chevauchent, de manière non ambiguë,

l'extrémité 3' de la séquence DAN1. Seules les séquences communes aux trois EST ont été ici ajoutées à la séquence DAN1. La chaîne polyA interne à la position 4002 diffère en longueur des trois EST (indiqués par un a en lettre minuscule) et n'est pas précédée par un signal de polyadénylation. EST N90240 présente cependant deux signaux de polyadénylation putatifs AATAAA situés à 33 et 59 paires de bases en 3' de l'extrémité de la séquence consensus proposée. Le premier codon méthionine à la position 243 et le consensus Kozak qui le précède sont soulignés. Ce codon est en phase avec une séquence d'acides aminés putative située en amont (en italique). Le premier codon de terminaison dans le cadre (position 2745) est souligné. Le cadre de lecture ouvert chevauchant est également montré (en italique) et son codon de terminaison souligné (position 3638). La séquence de la position 2560 à la position 2880 est confirmée par des EST chevauchant (H70616, R00491, R10603), ce qui écarte une mutation artéfactuelle du cadre de lecture dans le clone DAN1.

Un cadre de lecture ouvert commence à la position 1 jusqu'à la position 2745. La première méthionine est à la position 243 et est précédée d'une très bonne séquence consensus pour l'initiation de la traduction (accord de 6/9 y compris l'important A à -3). La séquence amont est très riche en GC, ce qui pourrait expliquer l'absence de codons d'arrêt dans les trois cadres de lecture.

De manière inattendue, un second cadre de lecture ouvert de 348 codons chevauche, dans un cadre différent, le plus grand cadre de lecture ouvert. La probabilité pour que cela soit le résultat du hasard est faible; cela suggère une mutation du cadre de lecture dans le clone originel DAN1. Une comparaison avec la séquence de 3 EST chevauchant cette région (positions

2560 à 2880) a cependant confirmé la séquence du clone DAN1 et la présence du codon d'arrêt prédit. L'existence d'un déphasage du cadre de lecture lors de la traduction reste une possibilité distincte, d'autant plus que le  
5 programme informatique GRAIL la donne au cadre de lecture ouvert 3' un score "excellent" pour sa capacité à coder des protéines. Une autre possibilité est l'existence d'épissages alternatifs produisant diverses formes d'ARNm, certaines étant porteuses d'un décalage du cadre  
10 de lecture (exemple: le gène FMR1 fragile X mental retardation 1). Trois autres EST chevauchant étendent la région 3' non traduite de 177 paires de bases (voir figure 7).

Un fragment DAN1 de 2,5 kb a été utilisé comme  
15 sonde dans des transferts Northern avec de l'ARN polyA + humain. Une expression ubiquiste a été trouvée dans différentes régions du cerveau et une forte expression a été observée dans d'autres organes.

En figure 8, est représenté une analyse d'un  
20 transfert Northern. Une sonde de 2,5 kb (de la position 1370 à 3985 sur la figure 7) a été utilisée pour les transferts Northern (MTN Clontech® et MTN2 issu de cerveau) contenant de l'ARN polyA + humain des régions de cerveau et des tissus indiqués. La longueur de l'ARNm a  
25 été évaluée à 4,4 kb, ce qui est très proche de la séquence de 4,2 kb dérivée du DAN1 et chevauchant les EST.

La ou les protéine(s) prédite(s) à partir des deux cadres de lecture ouverts ne présente(nt) pas  
30 d'homologie significative avec des protéines connues chez les humains ou chez les autres organismes. Le gène apparaît bien conservé chez les mammifères (bovins, lapins, moutons, cochons et souris) et les poulets comme l'indiquent les fragments à fortes réactions croisées  
35 observés sur transfert Zoo en utilisant les mêmes sonde

d'ADNc et conditions d'hybridation et de lavage que pour l'analyse par transfert Northern.

5

### . Discussion

SCA2 est le sixième locus cloné correspondant à une maladie où se trouve impliquée une extension de répétitions CAG/polyglutamine.

10 Dans la population étudiée, une limite inférieure de pathogénicité de 37 glutamines a été trouvée, ce qui est très proche de la limite de 36 trouvée chez les très rares patients atteints de la maladie de Huntington. Cette limite n'est pas définitive  
15 et ne pourra être établie que par l'étude d'un nombre plus élevé de patients. Les limites inférieures pour les autres maladies sont respectivement 40, 40, 49 et 61 pour SBMA, SCA1, DRPLA et MJD/SCA3.

20 Malgré un seuil de pathogénicité similaire, l'effet des glutamines additionnelles apparaît plus important dans SCA2 que dans HD puisque le déclenchement juvénile de la maladie (inférieur ou égal à 20 ans) est atteint avec 46 répétitions pour SCA2 alors que, pour HD, les cas juvéniles sont majoritairement atteints avec plus  
25 de 60 répétitions et, pour SCA1, à plus de 55 répétitions.

La courbe, très abrupte, de corrélation semble plus proche de celle de la maladie de Machado-Joseph (SCA3), pour laquelle le seuil de pathogénicité est  
30 cependant plus élevé. Cela suggère que la protéine impliquée dans SCA2 (ataxine 2) est très sensible aux polyglutamines.

En alternative, cette sensibilité accrue pourrait être une propriété des neurones affectés.  
35 L'existence d'un effet "contexte protéique" est supporté



par le fait que dans la protéine TBP (TATA binding protein), jusqu'à 42 glutamines sont trouvées dans les allèles normaux. Il est de plus possible que la toxicité d'une ataxine tronquée soit plus élevée que celle de la protéine entière, avec un rôle protecteur de la partie protéique tronquée.

Un autre caractère très frappant chez les familles SCA2 est la haute instabilité de la répétition. 13 des 16 transmissions mettent en évidence une instabilité et 5 d'entre elles, en particulier, présentent une augmentation de 5 à 10 répétitions.

De telles augmentations se produisent dans 20 à 30% des transmissions paternelles de HD, préférentiellement chez des allèles parentaux à plus de 45 glutamines, et sont très rares, pour SCA1 ou SBMA, pour des tailles de répétitions similaires.

Ces grandes extensions additionnelles se produisent, de plus, lors de transmissions paternelles et maternelles. Ce fait est en bonne corrélation avec le manque de biais parental dans l'anticipation de SCA2 pour laquelle les mêmes 11-16 années d'anticipation de l'âge de déclenchement de la maladie ont été observées indépendamment du sexe du parent transmetteur.

Le biais paternel pour l'extension qui est observée chez HD, SCA1, DRPLA et aussi chez des dystrophies myotoniques (pour des allèles dans la gamme des 50-100 CTG), et, dans une moindre mesure chez SBMA, n'est pas une propriété intrinsèque des répétitions CAG ou CTG. Des effets dus à la position (nature des séquences environnantes et localisation de la séquence relativement aux origines de réplication) pourraient jouer un rôle important dans l'instabilité.

Un autre trait du locus SCA2 est l'interruption, chez les allèles normaux, de la

répétition de CAG par 1 à 3 CAA (qui codent également pour des glutamines).

Les trois allèles pathologiques séquencés (issus de familles différentes) contiennent cependant des répétitions purement CAG. Ceci est très similaire au cas de SCA1 pour lequel les allèles normaux sont interrompus par des codons CAT (histidine) alors que les allèles étendus sont constitués purement de CAG.

Ces interruptions dans les répétitions pourraient avoir un effet stabilisant et la perte des interruptions CAA pourrait constituer un événement initial dans l'histoire de la mutation.

Comme dans quatre des cinq autres maladies à polyglutamine (l'exception étant SBMA et le gène du récepteur androgène), le gène SCA2 n'a pas de fonction évidente et apparaît comme s'exprimant de manière ubiquiste dans le cerveau et cela même dans des zones telles que le putamen qui reste non affecté chez les patients.

La taille apparente de la protéine SCA2 mutante est de 150 kDa sur transferts Western mais les chaînes étendues de polyglutamines peuvent affecter la migration électrophorétique et une taille d'environ 120-130 kDa (environ 1100-1200 acides aminés) pour l'ataxine 2 normale peut être estimée.

Le cadre de lecture ouvert principal commençant au premier codon méthionine (caractérisé par une très bonne concordance avec le consensus Kozak), code pour une protéine de 834 acides aminés et d'un poids moléculaire de 89,9 kDa.

Pour expliquer un tel écart, on peut supposer que, soit le vrai codon d'initiation est en amont de la séquence présentée, soit le cadre de lecture ouvert 3' qui chevauche le cadre principal est utilisé par déphasage du cadre de lecture lors de la traduction sur

les ribosomes. Une troisième possibilité est que les ADN  
séquencés correspondent à un épissage alternatif altérant  
la phrase de lecture. Etant donné que la séquence  
présentée (4,2 kb) est très proche de la longueur d'ARNm  
5 estimée sur transfert Northern (4,4 kb, polyA inclus), il  
n'y a pas beaucoup de place pour un site d'initiation  
amont additionnel.

Il est donc possible qu'un déphasage  
ribosomique du cadre de lecture, qui est observé pour  
10 plusieurs gènes viraux et aussi pour quelques gènes  
humains, soit impliqué.

Une mutation -1 du cadre de lecture  
correspondrait à une protéine normale de 1132 acides  
aminés et de poids moléculaire 121,7 kDa, ce qui est en  
15 très bon accord avec le poids moléculaire estimé.

Par analyse informatique de la séquence ARNm,  
des structures secondaires de pseudonoeuds, qui peuvent  
constituer des séquences stimulant les événements de  
mutation du cadre de lecture, n'ont pu être mises en  
20 évidence. La mutation -1 du cadre de lecture telle que  
suggérée ci-dessus ne constitue donc qu'une explication  
possible des écarts de tailles observés.

La mise en évidence d'un éventuel épissage  
alternatif pourra être établie par RT-PCR en utilisant  
25 des amorces basées sur la séquence de la figure 7 à  
partir d'ARN provenant de divers tissus.

## LISTE DE SEQUENCES

## (1) INFORMATIONS GENERALES:

## (i) DEPOSANT:

(A) NOM: C.N.R.S.  
 (B) RUE: 3, rue Michel-Ange  
 (C) VILLE: PARIS  
 (E) PAYS: FRANCE  
 (F) CODE POSTAL: 75016

(A) NOM: I.N.S.E.R.M.  
 (B) RUE: 101, rue de Tolbiac  
 (C) VILLE: PARIS  
 (E) PAYS: FRANCE  
 (F) CODE POSTAL: 75013

(ii) TITRE DE L' INVENTION: Moyens pour le traitement et le diagnostic des maladies neurodégénératives

(iii) NOMBRE DE SEQUENCES: 6

## (iv) FORME DECHIFFRABLE PAR ORDINATEUR:

(A) TYPE DE SUPPORT: Floppy disk  
 (B) ORDINATEUR: IBM PC compatible  
 (C) SYSTEME D' EXPLOITATION: PC-DOS/MS-DOS  
 (D) LOGICIEL: PatentIn Release #1.0, Version #1.30 (OEB)

## (2) INFORMATIONS POUR LA SEQ ID NO: 1:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 399 paires de bases  
 (B) TYPE: nucleotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: des deux

(ii) TYPE DE MOLECULE: ADNC

(iii) HYPOTHETIQUE: NON

(iv) ANTI-SENS: NON

(vii) SOURCE IMMEDIATE:

(B) CLONE: AAD10

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 1:

CAGCAGCAGC	AGCAGCAGCA	ACAGCAGCAG	CAGNNNNNNN	40
NNNNNNNNNN	NNNNNNNCAG	CAGCAACAGC	AGCAGNNNNN	80
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	120
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	160
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	200
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	240
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	280
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	360
NNNNNNNCAGC	AGCAGCAGCA	GCAACAACAG	CACCAACAGC	400
AA				402

## (3) INFORMATIONS POUR LA SEQ ID NO: 2:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 152 paires de bases
- (B) TYPE: nucleotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: des deux

(ii) TYPE DE MOLECULE: ADNc

(iii) HYPOTHETIQUE: NON

(iv) ANTI-SENS: NON

## (vii) SOURCE IMMEDIATE:

(B) CLONE: AAD14

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 2:

CAGCAGCAGC AGCAGCAGCA GCAACAGCAG CAGCAGNNNN	40
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	80
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NCAGCAGCAG	120
CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG CAG	153

## (4) INFORMATIONS POUR LA SEQ ID NO: 3:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 66 paires de bases
- (B) TYPE: nucleotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: des deux

(ii) TYPE DE MOLECULE: ADNc

(iii) HYPOTHETIQUE: NON

(iv) ANTI-SENS: NON

## (vii) SOURCE IMMEDIATE:

(B) CLONE: DAN1

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 3:

CAGCAGCAGC AGCAGCAGCA GCAGCAACAG CAGCAGCAGC	40
AACAGCAGCA GCAGCAGCAG CAGCAG	66

## (5) INFORMATIONS POUR LA SEQ ID NO: 4:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 54 paires de bases
- (B) TYPE: nucleotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: des deux

- (ii) TYPE DE MOLECULE: ADNc
- (iii) HYPOTHETIQUE: NON
- (iv) ANTI-SENS: NON
- (vii) SOURCE IMMEDIATE:
  - (B) CLONE: DAN15

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 4:

CAGCAGCAGC AGCAGCAGCA GCAACAGCAG CAGCAGCAGC 40  
AGCAGCAACA GCAA 54

(6) INFORMATIONS POUR LA SEQ ID NO: 5:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
  - (A) LONGUEUR: 36 paires de bases
  - (B) TYPE: nucleotide
  - (C) NOMBRE DE BRINS: simple
  - (D) CONFIGURATION: des deux
- (ii) TYPE DE MOLECULE: ADNc
- (iii) HYPOTHETIQUE: NON
- (iv) ANTI-SENS: NON
- (vii) SOURCE IMMEDIATE:
  - (B) CLONE: DAN26

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 5:

CAGCAGCAGC AGCAGCAGCA ACAGCAGCAG CAGCAG 36

(7) INFORMATIONS POUR LA SEQ ID NO: 6:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
  - (A) LONGUEUR: 42 paires de bases
  - (B) TYPE: nucleotide
  - (C) NOMBRE DE BRINS: simple
  - (D) CONFIGURATION: des deux
- (ii) TYPE DE MOLECULE: ADNc
- (iii) HYPOTHETIQUE: NON
- (iv) ANTI-SENS: NON
- (vii) SOURCE IMMEDIATE:
  - (B) CLONE: AAD20

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 6:

CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AG 42

## (8) INFORMATIONS POUR LA SEQ ID NO: 7:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 4199 paires de bases
- (B) TYPE: nucleotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: des deux

(ii) TYPE DE MOLECULE: ADNc

(iii) HYPOTHETIQUE: NON

(iv) ANTI-SENS: NON

(vii) SOURCE IMMEDIATE:

(B) CLONE: DAN1

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 1:

ACGGCAACGG CGGCGGCGCG TTTCGGCCCG GCTCCCGGCG GTCCTTGGT CTCGGCGGGC	60
CTCCCCGCCC CTTCGTCGTC GTCCTTCTCC CCCTCGCCAG CCGGGGCGCC CCTCCGGCCG	120
CGCCAACCCG CGCCTCCCCG CTCGGCGCCC GTGCGTCCCC GCCGCGTTCC GGCGTCTCCT	180
TGGCGCGCCC GGCTCCCGGC TGTCCCCGCC CGGCGTGCGA GCCGGTGTAT GGGCCCCTCA	240
CCATGTCGCT GAAGCCCCAG CAGCAGCAGC AGCAGCAGCA GCAACAGCAG CAGCAGCAAC	300
AGCAGCAGCA GCAGCAGCAG CAGCCGCCGC CCGCGGCTGC CAATGTCCGC AAGCCCGGCG	360
GCAGCGGCCT TCTAGCGTCG CCCGCCGCCG CGCCTTCGCC GTCCTCGTCC TCGGTCTCCT	420
CGTCTCGGC CACGGCTCCC TCCTCGGTGG TCGCGGCGAC CTCCGGCGGC GGGAGGCCCG	480
GCCTGGGCAG AGGTCGAAAC AGTAACAAAG GACTGCCTCA GTCTACGATT TCTTTTGATG	540
GAATCTATGC AAATATGAGG ATGGTTCATA TACTTACATC AGTTGTTGGC TCCAAATGTG	600
AAGTACAAGT GAAAAATGGA GGTATATATG AAGGAGTTTT TAAACTTAC AGTCCGAAGT	660
GTGATTTGGT ACTTGATGCC GCACATGAGA AAAGTACAGA ATCCAGTTCG GGGCCGAAAC	720
GTGAAGAAAT AATGGAGAGT ATTTTGTTC AATGTTTACA CTTTGTTGTG GTACAGTTTA	780
AAGATATGGA CTCCAGTTAT GCAAAAAGAG ATGCTTTTAC TGA CTCTGCT ATCAGTGCTA	840
AAGTGAATGG CGAACACAAA GAGAAGGACC TGGAGCCCTG GGATGCAGGT GAACTCACAG	900
CCAATGAGGA ACTTGAGGCT TTGGAAAATG ACGTATCTAA TGGATGGGAT CCCAATGATA	960
TGTTTCGATA TAATGAAGAA AATTATGGTG TAGTGTCTAC GTATGATAGC AGTTTATCTT	1020
CGTATACAGT GCCCTTAGAA AGAGATAACT CAGAAGAATT TTAAAACGG GAAGCAAGGG	1080
CAAACCAAGT AGCAGAAGAA ATTGAGTCAA GTGCCAGTA CAAAGCTCGA GTGGCCCTGG	1140
AAAATGATGA TAGGAGTGAG GAAGAAAAAT ACACAGCAGT TCAGAGAAAT TCCAGTGAAC	1200
GTGAGGGGCA CAGCATAAAC ACTAGGGAAA ATAAATATAT TCCTCCTGGA CAAAGAAATA	1260

GAGAAGTCAT ATCCTGGGGA AGTGGGAGAC AGAATTCACC GCGTATGGGC CAGCCTGGAT 1320  
CGGGCTCCAT GCCATCAAGA TCCACTTCTC ACACTTCAGA TTTCAACCCG AATTCTGGTT 1380  
CAGACCAAAG AGTAGTTAAT GGAGGTGTTC CCTGGCCATC GCCTTGCCCA TCTCCTTCCT 1440  
CTCGCCACCC TTCTCGCTAC CAGTCAGGTC CCAACTCTCT TCCACCTCGG GCAGCCACCC 1500  
CTACACGGCC GCCCTCCAGG CCCCCCTCGC GGCCATCCAG ACCCCCGTCT CACCCCTCTG 1560  
CTCATGGTTC TCCAGCTCCT GTCTCTACTA TGCCTAAACG CATGTCTTCA GAAGGGCCTC 1620  
CAAGGATGTC CCCAAAGGCC CAGCGACATC CTCGAAATCA CAGAGTTTCT GCTGGGAGGG 1680  
GTTCCATATC CAGTGGCCTA GAATTTGTAT CCCACAACCC ACCCAGTGAA GCAGCTACTC 1740  
CTCCAGTAGC AAGGACCAGT CCCTCGGGGG GAACGTGGTC ATCAGTGGTC AGTGGGGTTC 1800  
CAAGATTATC CCCTAAACT CATAGACCCA GGTCTCCCAG ACAGAACAGT ATTGGAAATA 1860  
CCCCCAGTGG GCCAGTTCTT GCTTCTCCCC AAGCTGGTAT TATTCCAAT GAAGCTGTTG 1920  
CCATGCCTAT TCCAGCTGCA TCTCCTACGC CTGCTAGTCC TGCATCGAAC AGAGCTGTTA 1980  
CCCCTTCTAG TGAGGCTAAA GATTCCAGGC TTCAAGATCA GAGGCAGAAC TCTCCTGCAG 2040  
GGAATAAAGA AAATATTAAA CCCAATGAAA CATCACCTAG CTTCTCAAAA GCTGAAAACA 2100  
AAGGTATATC ACCAGTTGTT TCTGAACATA GAAAACAGAT TGATGATTTA AAGAAATTTA 2160  
AGAATGATTT TAGGTTACAG CCAAGTTCTA CTTCTGAATC TATGGATCAA CTAATAACA 2220  
AAAATAGAGA GGGAGAAAAA TCAAGAGATT TGATCAAAGA CAAAATTGAA CCAAGTGCTA 2280  
AGGATTCTTT CATTGAAAAT AGCAGCAGCA ACTGTACCAG TGGCAGCAGC AAGCCGAATA 2340  
GCCCCAGCAT TTCCCCTTCA ATACTTAGTA ACACGGAGCA CAAGAGGGGA CCTGAGGTCA 2400  
CTTCCCAAGG GGTTCAGACT TCCAGCCCAG CATGTAAACA AGAGAAAGAC GATAAGGAAG 2460  
AGAAGAAAGA CGCAGCTGAG CAAGTTAGGA AATCAACATT GAATCCCAAT GCAAAGGAGT 2520  
TCAACCCACG TTCCTTCTCT CAGCCAAAGC CTTCTACTAC CCCAACTTCA CCTCGGCCTC 2580  
AAGCACAACC TAGCCCATCT ATGGTGGGTC ATCAACAGCC AACTCCAGTT TATACTCAGC 2640  
CTGTTTGTTT TGCACCAAAT ATGATGTATC CAGTCCCAGT GAGCCCAGGC GTGCAATACC 2700  
AAATATGCCC CAACAGCGGC AAGACCAGCA TCATCAGAGT GCCATGATGC ACCCAGCGTC 2760  
AGCAGCGGGC CCACCGATTG CAGCCACCCC ACCAGCTTAC TCCAGCAAT ATGTTGCCTA 2820  
CAGTCCTCAG CAGTTCCCAA ATCAGCCCCT TGTTCAAGAT GTGCCACATT ATCAGTCTCA 2880  
GCATCCTCAT GTCTATAGTC CTGTAATACA GGGTAATGCT AGAATGATGG CACCACCAAC 2940  
ACACGCCCAG CCTGGTTTAG TATCTTCTTC AGCAACTCAG TACGGGGCTC ATGAGCAGAC 3000  
GCATGCGATG TATGCATGTC CCAAATTACC ATACAACAAG GAGACAAGCC CTTCTTTCTA 3060  
CTTTGCCATT TCCACGGGCT CCCTTGCTCA GCAGTATGCG CACCCTAACG CTACCCTGCA 3120  
CCCACATACT CCACACCCTC AGCCTTCAGC TACCCCCACT GGACAGCAGC AAAGCCAACA 3180



TGGTGGAAGT CATCCTGCAC CCAGTCCTGT TCAGCACCAT CAGCACCAGG CCGCCCAGGC	3240
TCTCCATCTG GCCAGTCCAC AGCAGCAGTC AGCCATTTAC CACGCGGGGC TTGCGCCAAC	3300
TCCACCCTCC ATGACACCTG CCTCCAACAC GCAGTCGCCA CAGAATAGTT TCCCAGCAGC	3360
ACAACAGACT GTCTTTACGA TCCATCCTTC TCACGTTTCT CCGGCGTATA CCAACCCACC	3420
CCACATGGCC CACGTACCTC AGGCTCATGT ACAGTCAGGA ATGGTTTCCTT CTCATCCAAC	3480
TGCCCATGCG CCAATGATGC TAATGACGAC ACAGCCACCC GCGGGTCCCC AGGCCGCCCT	3540
CGCTCAAAGTG CACTACAGCC CATTCCAGTC TCGACAACAG CGCATTTCCTT CTATATGACG	3600
CACCCTTCAG TACAAGCCCA CCACCAACAG CAGTTGTAAG GCTGCCCTGG AGGAACCGAA	3660
AGGCCAAATT CCCTCCTCCC TTCTACTGCT TCTACCAACT GGAAGCACAG AAAACTAGAA	3720
TTTCATTTAT TTTGTTTTTA AAATATATAT GTTGATTTCT TGTAACATCC AATAGGAATG	3780
CTAACAGTTC ACTTGAGTG GAAGATACTT GGACCGAGTA GAGGCATTTA GGAACCTGGG	3840
GGCTATTCCA TAATTCCATA TGCTGTTTCA GAGTCCCGCA GGTACCCAG CTCTGCTTGC	3900
CGAAACTGGA AGTTATTTAT TTTTAATAA CCCTTGAAAG TCATGAACAC ATCAGCTAGC	3960
AAAAGAAGTA ACAAGAGTGA TTCTTGCTGC TATTACTGCT AAAAAAAAAA AAAAAAAAAA	4020
AaaaaaaaaTC AAGACTTGGA ACGCCCTTTT ACTAACTTG ACAAAGTTTC AGTAAATTCT	4080
TACCGTCAAA CTGACGGATT ATTATTTATA AATCAAGTTT GATGAGGTGA TCACTGTCTA	4140
CAGTGGTTCA ACTTTTAAGT TAAGGGAAAA ACTTTTACTT TGTAGATAAT ATAAAATCC	4199

## REVENDEICATIONS

1. Utilisation de l'anticorps 1C2 ou d'un fragment ou d'un dérivé de l'anticorps 1C2 pour la  
5 préparation d'une composition pharmaceutique destinée au traitement préventif ou curatif des maladies neurodégénératives associées à une répétition de glutamine.
2. Utilisation selon la revendication 1  
10 caractérisée en ce qu'elle met en oeuvre l'anticorps 1C2.
3. Utilisation selon la revendication 1 caractérisée en ce qu'elle met en oeuvre un fragment ScFv de l'anticorps 1C2.
4. Utilisation d'un acide nucléique codant pour  
15 un fragment ScFv de l'anticorps 1C2 pour la préparation d'une composition pharmaceutique destinée au traitement préventif ou curatif des maladies neurodégénératives associées à une répétition de glutamine.
5. Utilisation selon l'une quelconque des  
20 revendications 1 à 4 pour le traitement des maladies telles que par exemple la maladie de Huntington, l'ataxie spinocérébelleuse de type 1, 2, 3, 4, 5 ou 7, l'atrophie musculaire spino-bulbaire associée au chromosome X ou maladie de Kennedy, l'atrophie dentarorubral-  
25 pallidoluysienne, l'ataxie spinocérébelleuse autosomale dominante, et la paraplégie spastique familiale, ou encore la maladie affective bipolaire, la psychose maniaco-dépressive ou la schizophrénie.
6. Molécule comprenant le site de liaison de la  
30 région variable de la chaîne légère de l'anticorps 1C2 relié par un bras peptidique au site de liaison de la région variable de la chaîne lourde de l'anticorps 1C2.
7. Séquence d'acide nucléique caractérisée en  
35 ce qu'elle code pour la molécule selon la revendication 6.

8. Vecteur comprenant la séquence selon la revendication 7 sous contrôle d'un promoteur fonctionnel dans les cellules de mammifères.

5 9. Vecteur selon la revendication 8 caractérisé en ce qu'il s'agit d'un adénovirus recombinant défectif.

10. Vecteur selon la revendication 8 caractérisé en ce qu'il s'agit d'un rétrovirus recombinant défectif.

10 11. Vecteur selon la revendication 8 caractérisé en ce qu'il s'agit d'un virus adéno associé recombinant défectif.

12. Vecteur selon la revendication 8 caractérisé en ce qu'il s'agit du virus de l'herpès recombinant défectif.

15 13. Composition pharmaceutique comprenant l'anticorps 1C2, un fragment ou dérivé de cet anticorps.

14. Composition pharmaceutique comprenant un vecteur selon l'une quelconque des revendications 8 à 12.

20 15. Composition pharmaceutique destinée au traitement des maladies neurodégénératives associées à la présence d'une protéine portant une chaîne homopolymère de glutamine caractérisée en ce qu'elle comprend l'anticorps 1C2, un fragment ou dérivé de cet anticorps, notamment selon la revendication 6, ou un vecteur selon  
25 l'une quelconque des revendications 8 à 12.

30 16. Méthode de diagnostic des maladies neurodégénératives caractérisée en ce que l'on détecte in vitro, au moyen de l'anticorps 1C2, a un fragment ou d'un dérivé de l'anticorps 1C2 la présence de protéines portant une chaîne polyglutamine de longueur pathologique.

17. Méthode selon revendication 16 caractérisée en ce que la détection de ces protéines est réalisée par mise en contact d'un extrait cellulaire avec l'anticorps

monoclonal 1C2 et révélation des complexes anticorps-antigènes formés.

18. Méthode selon l'une des revendications 16 ou 17 caractérisée en ce que la détection est réalisée  
5 sur un extrait de cellules sanguines.

19. Méthode selon l'une quelconque des revendications 16 à 18 pour la mise en évidence de prédisposition à la maladie de Huntington ou à une ataxie spinocérébelleuse 1, 2 ou 3.

10 20. Méthode selon l'une quelconque des revendications précédentes 16 à 18 pour la mise en évidence de maladies dégénératives du système nerveux central causées par la présence d'une chaîne polyglutamine dans une protéine exprimée.

15 21. ADN caractérisé en ce qu'il comporte tout ou partie d'au moins une des séquences SEQ ID n°1, n°2, n°3, n°4, n°5, n°6, n°7 ou d'au moins une séquence présentant une homologie supérieure ou égale à 50% avec ces séquences.

20 22. Les ARN correspondant à la transcription d'au moins un ADN selon la revendication 21, les séquences complémentaires de ces ADN ou ARN, ou leurs séquences anti-sens.

25 23. Procédé de criblage d'expression d'ADNs à répétition CAG ou de leurs produits d'expression caractérisé en ce qu'il comprend l'utilisation de l'anticorps 1C2 ou d'un fragment ou d'un dérivé de cet anticorps.

30 24. Procédé d'identification ou de purification de protéines à chaînes polyglutaminiques utilisant une étape d'immunodétection ou d'immunopurification par l'anticorps 1C2, fragment ou dérivé de cet anticorps, ou pouvant conduire secondairement à identifier le gène correspondant.

25. Méthode de diagnostic utilisant  
l'amplification PCR sur ADN ou RT-PCR sur ARN permettant  
de détecter des formes mutées dans des gènes codant pour  
des chaînes polyglutaminiques identifiées ou clonées  
5 grâce à l'anticorps 1C2.

10

15

longueur



Figure 3

2/8

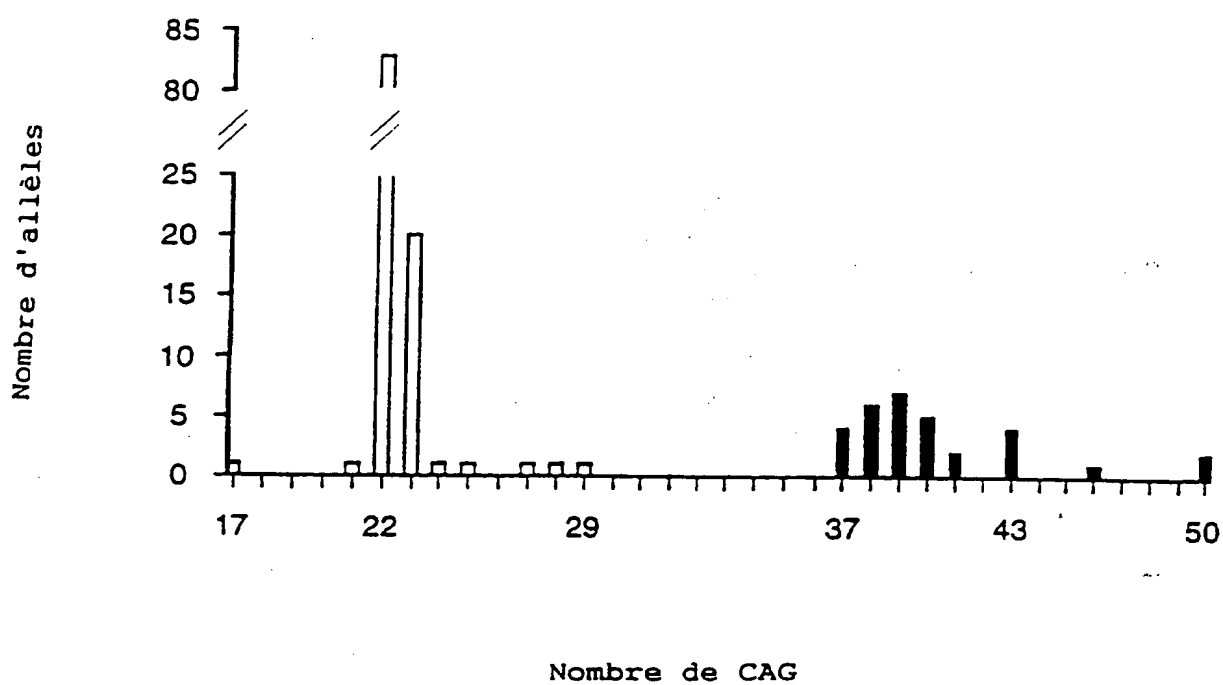


Figure 4

3/8

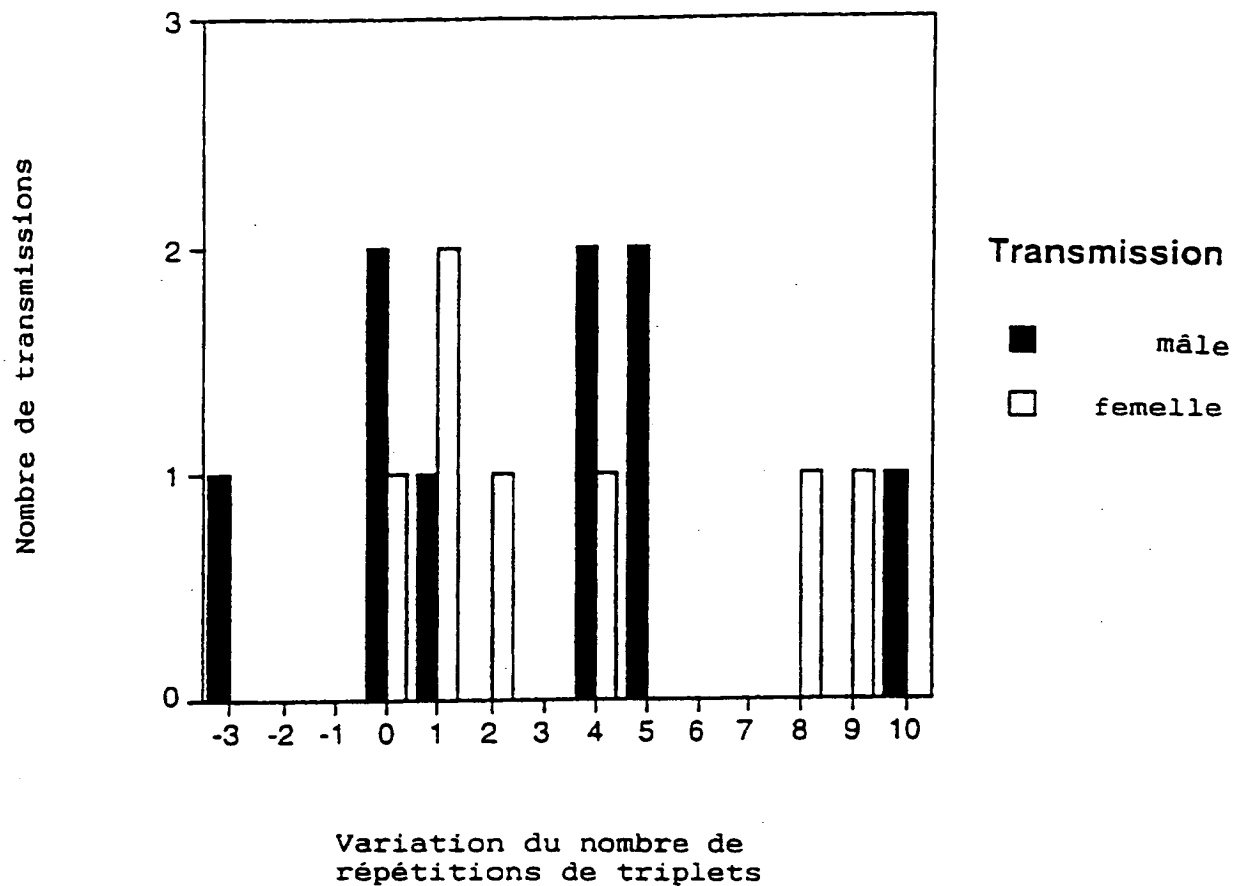


figure 5



4/8

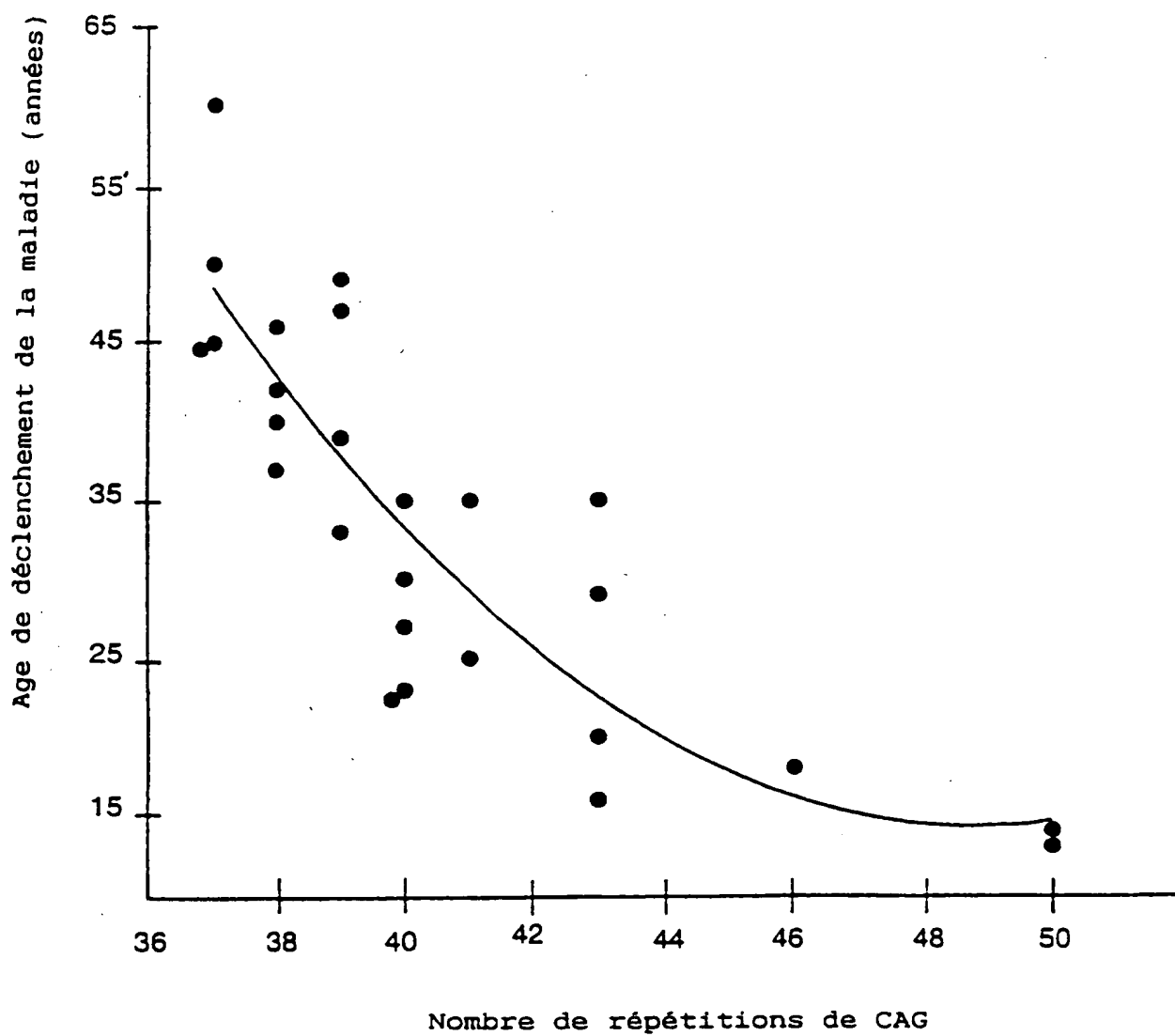


Figure 6

figure 7

[illegible]

## figure 7 (suite 1)

1725 JTGAGCCAGCTACTCTCCAGTAGCAAGGACCAAGTCCCTCCGGGGGAACCTGCTTATTAGTGGTCAGTGGGGTTC 1300  
E A A T P P V A R T S P S G G T W S S V V S G V P

1801 CAAGATTATCCCTAAAACTCATAGACCCAGGTCTCCAGACAGAACAGTATTGGAAATACCCCCAGTGGGCCAG 1875  
R L S P K T H R P R S P R Q N S I G N T P S G P V

1876 TTCTTGGCTTCTCCCAAGCTGGTATTATTCCAACTGAAGCTGTTGCCATGCCATTCCAGCTGCATCTCTACGC 1950  
L A S P Q A G I I P T E A V A M P I P A A S P T P

1951 CTGCTAGTCTGCAATCGAACAGAGCTGTTACCCCTTCTAGTGAGGCTAAAGATTCCAGGCTTCAAGATCAGAGGC 2025  
A S P A S N R A V T P S S E A K D S R L Q D Q R Q

2026 AGAACTCTCTGCGAGGGAATAAAGAAAATATTAAACCAATGAACATCACCTAGCTTCTCAAAAGCTGAAAACA 2100  
N S P A G N K E N I K P N E T S P S F S K A E N K

2101 AAGGTATATCACCAGTTGTTTCTGAACATAGAAAACAGATTGATGATTTAAAGAAATTTAAGAATGATTTTAGGT 2175  
G I S P V V S E H R K Q I D D L K K F K N D F R L

2176 TACAGCCAAGTTCTACTTCTGAATCTATGGATCAACTACTAAACAAAAATAGAGAGGGAGAAAAATCAAGAGATT 2250  
Q P S S T S E S M D Q L L N K N R E G E K S R D L

2251 TGATCAAGACAAAATTGAACCAAGTGCTAAGGATTCTTTCAATTGAAAATAGCAGCAGCAACTGTACCAGTGGCA 2325  
I K D K I E P S A K D S F I E N S S S N C T S G S

2326 GCAGCAAGCCGAATAGCCCCAGCATTTCCTTCAATCTTAGTAACACGGAGCACAAGAGGGGACCTGAGGTCA 2400  
S K P N S P S I S P S I L S N T E H K R G P E V T

2401 CTTCCCAAGGGGTTCCAGACTTCCAGCCAGCATGTAAACAAGAGAAAGACGATAAGGAAGAGAAGAAAGACGCAG 2475  
S Q G V Q T S S P A C K Q E K D D K E E K K D A A

2476 CTGAGCAAGTTAGGAAATCAACATTGAATCCCAATGCAAAGGAGTTCAACCCACGTTCTCTCTCAGCCAAAGC 2550  
E Q V R K S T L N P N A K E F N P R S P S Q P K P

2551 CTTCTACTACCCCACTTCACCTCGGCCCTCAAGCACAACCTAGCCCATCTATGGTGGGTCAACACAGCCAACTC 2625  
S T T P T S P R P Q A Q P S P S M V G H Q Q P T P  
P I Y G G S S T A N S

2626 CAGTTTATACTCAGCCTGTTTGTGTTGTCACCAATATGATGTATCCAGTCCCAGTGAGCCAGGGCGTGAATACC 2700  
V Y T Q P V C F A P N M M Y P V P V S P G V Q Y Q  
S L Y S A C L P C T K Y D V S S P S E P R R A I P

2701 AAATATGCCCCAACAGCGGCAAGACCAGCATCATCAGAGTGCCATGATGACCCAGCGTCAGCAGCGGGCCCCACC 2775  
I C P N S G K T S I I R V P  
N M P Q Q R Q D Q H H Q S A M M H P A S A A G P P

2776 GATTGCAGCCACCCACAGCTTACTCCACCGAATATGTTGCCATCAGTCTCAGCAGTTCCCAAATCAGCCCTT 2850  
I A A T P P A Y S T Q Y V A Y S P Q Q P P N Q P L

2851 TGTTCCAGCATGTGCCACATTATCAGTCTCAGCATCTCTCATGTCTATAGTCTGTAAATACAGGGTAATGCTAGAAT 2925  
V Q H V P H Y Q S Q H P H V Y S P V I Q G N A R M

2926 GATGGCACCACCAACACAGCCCGAGCTGGTTTAGTATCTTCTTCAGCAACTCAGTACGGGGCTCATGAGCAGAC 3000  
H A P P T H A Q P G L V S S S A T Q Y G A H E Q T

3001 GCATGCGATGTATGCATGTCCCAATTACCATACAACAAGGAGACAAGCCCTTCTTTCTACTTTGCCATTTCCAC 3075  
H A M Y A C P K L P Y N K E T S P S F Y F A I S T

3076 GGGCTCCCTTGCTCAGCAGTATGCGCACCCCTAACGCTACCCCTGCACCCACATACTCCACACCCCTCAGCCTTCAGC 3150  
G S L A Q Q Y A H P N A T L H P H T P H P Q P S A

3151 TACCCCCACTGGACAGCAGCAAGCCAACATGGTGGAGTCACTCTGCACCCAGTCTGTTCAGCACCATCAGCA 3225  
T P T G Q Q Q S Q H G G S H P A P S P V Q H H Q H

3226 CCAGGCGCGCCAGGCTCTCCATCTGGCCAGTCCACAGCAGCAGTCAGCCATTTACCACGCGGGGCTTGCSCCAAC 3300  
Q A A Q A L H L A S P Q Q Q S A I Y H A G L A P T

3301 TTCACCCCTCATGACACCTGCGCTCCAACAGCGAGTCGCCACAGAATAGTTTCCAGCAGCACAACAGACTGTCTT 3375  
P P S M T P A S N T Q S P Q N S F P A A Q Q T V F

## figure 7 (suite 2 et fin)

3375 TACGATCCATCCTTCTCAGCTTCAGCCGGCGTATACCAACCCACCCACATGCGCCACGTACCTCAGGCTCATGT 3450  
T I H P S H V Q P A Y T N P P H M A H Y P Q A H V

3451 ACAGTCAGGAATGGTTCTCTCATCCAACTGCCCATGCGCCAATGATGCTAATGACGACACAGCCACCCGGCGG 3525  
Q S G M V P S H P T A H A P M M L M T T Q P P G G

3526 TCCCCAGGCGCGCCTCGCTCAAAGTGCACCTACAGCCCATTCAGTCTCGACAACAGCGCATTTCCCTTATATGAC 3600  
P Q A A L A Q S A L Q P I P V S T T A H F P Y M T

3601 GCACCCCTTCAGTACAAGCCCAACCAACAGCAGTTGTAAAGGCTGCCCTGGAGGAACCGAAAGGCCAAATTCCTT 3675  
H P S V Q A H H Q Q Q L

3676 CCTCCCTTCTACTGCTTCTACCAACTGGAAGCACAGAAAACCTAGAATTTCAATTTATTTTGTATTTAAAAATATATA 3750

3751 TGTGATTTCTTGTAAATCCAAATAGGAATGCTAACAGTTCACTTGCAGTGGAAAGATACTTGGACCGAGTAGAGG 3825

3826 CATTTAGGAACTTGGGGGCTATTCCATAAATCCATATGCTGTTTCAGAGTCCCGCAGGTACCCAGCTCTGCTTG 3900

3901 CCGAAACTGGAAGTTATTTATTTTTTAATAACCCCTTGAAAGTCATGAACACATCAGCTAGCAAAAGAAAGTAACAA 3975

3976 GAGTGATTCTTCTGCTATTACTGCTAAAAAAAAAAAAAAAAAAAAAaTCAAGACTTGGAACGCTCTTT 4050

4051 TACTAACTTGACAAAGTTTCAGTAAATCTTACCCTCAAACCTGACGGATTATTTATTTATAAATCAAGTTTGATG 4125

4126 AGGTGATCACTCTCTACAGTGGTTCAACTTTTAAGTTA 4163

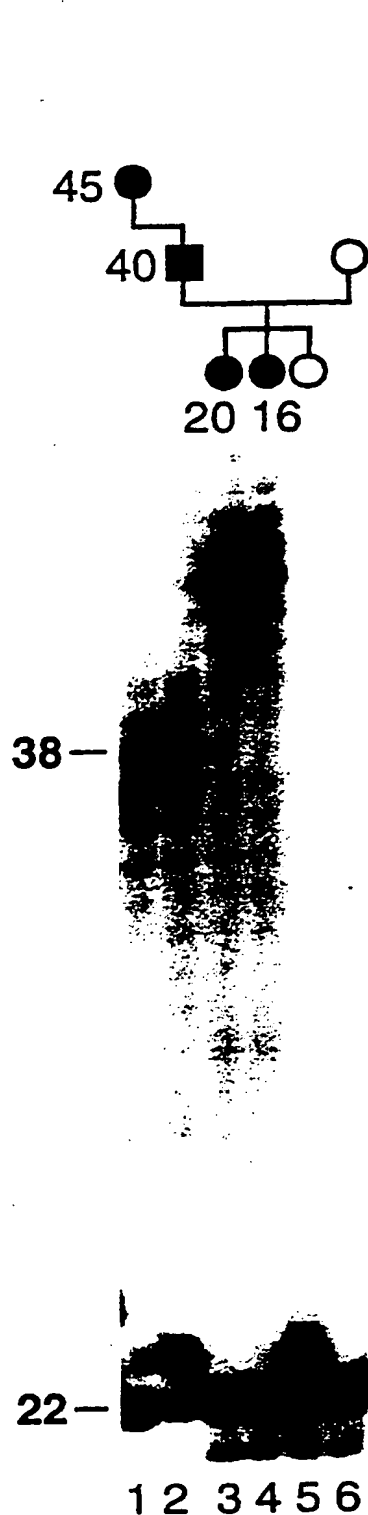


FIGURE 2

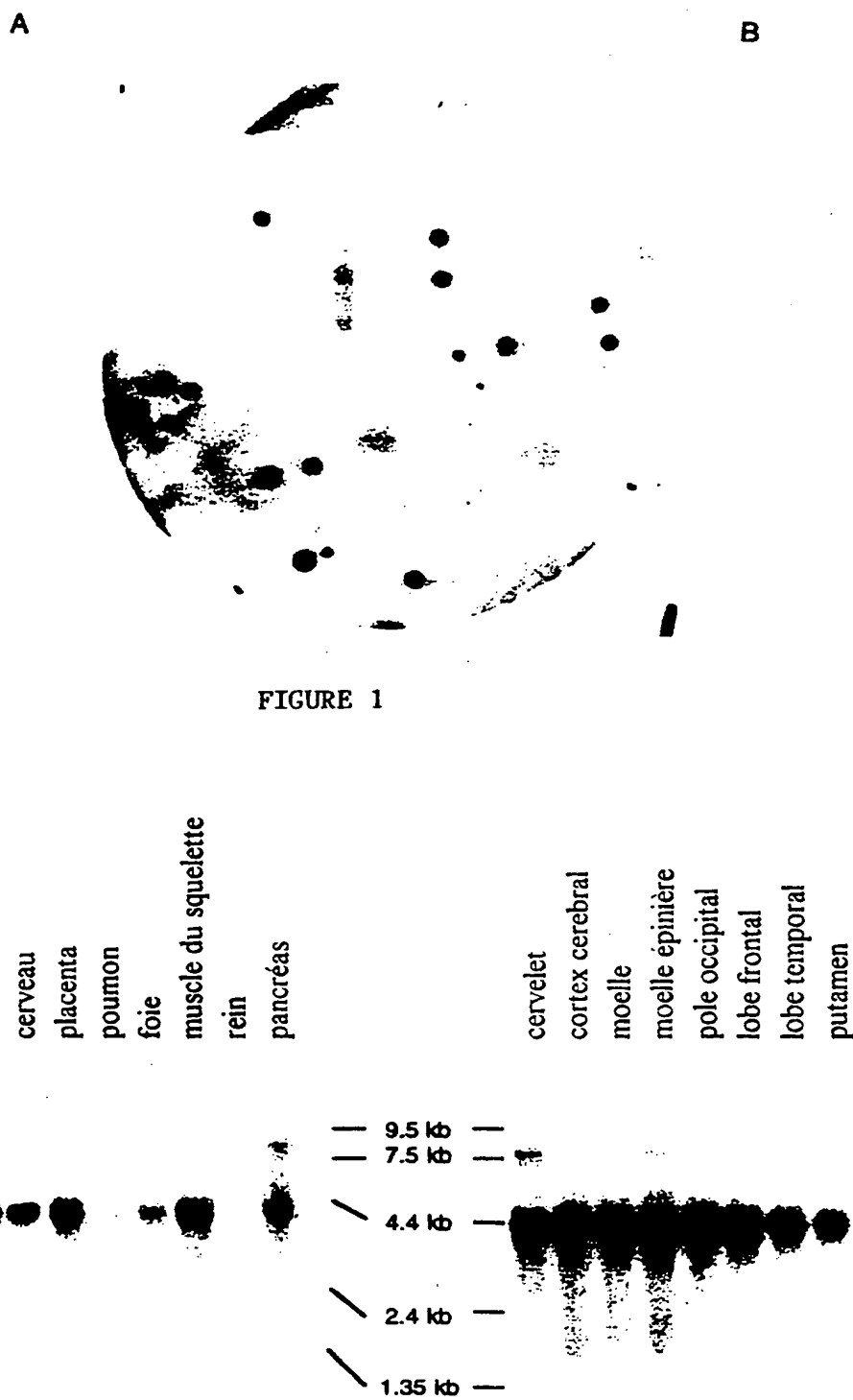


FIGURE 1

FIGURE 8

FEUILLE DE REMPLACEMENT (REGLE 26)

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/FR 96/01773

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/13 A61K39/395 A61K48/00 C07K16/18 C12N15/86  
G01N33/577 G01N33/68 C12Q1/68 C12N15/12

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A61K C07K G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data-base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 95 01437 A (REGENTS OF THE UNIVERSITY OF MINNESOTA) 12 January 1995 see example IX see claims	21,22
X	THE EMBO JOURNAL, vol. 13, no. 5, 1 March 1994, OXFORD, GRANDE BRETAGNE, pages 1166-1175, XP002009616 A. LESCURE ET AL.: "The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters." cited in the application see abstract see figure 1	23,24



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

### \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- \*A\* document member of the same patent family

Date of the actual completion of the international search

14 March 1997

Date of mailing of the international search report

24.03.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+ 31-70) 340-3016

Authorized officer

Nooij, F

# INTERNATIONAL SEARCH REPORT

Inter-  
national Application No  
PCT/FR 96/01773

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>NATURE GENETICS, vol. 10, no. 1, May 1995, NEW YORK, NY, ATATS-UNIS, pages 104-110, XP000577139 Y. TROTTIER ET AL.: "Cellular localization of the Huntington's disease protein and discrimination of the normal and mutated form." see the whole document</p> <p style="text-align: center;">---</p>	<p>1,2,5, 13, 15-20, 23-25</p>
A	<p>HUMAN MOLECULAR GENETICS, vol. 4, no. 3, March 1995, OXFORD, GRANDE BRETAGNE, pages 465-469, XP000577137 Y. JOU ET AL.: "Evidence from antibody studies that the CAG repeat in the Huntington disease gene is expressed in the protein." see abstract</p> <p style="text-align: center;">---</p>	<p>1,2,5, 13, 15-20, 23-25</p>
A	<p>BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 209, no. 3, 26 April 1995, DULUTH, MN, ATATS-UNIS, pages 1119-1125, XP002009615 K. IDE ET AL.: "Abnormal gene product identified in Huntington's disease lymphocytes and brain." see the whole document</p> <p style="text-align: center;">---</p>	<p>1,2,5, 13, 15-20, 23-25</p>
P,X	<p>NATURE, vol. 378, no. 6555, 23 November 1995, LONDRES, GRANDE BRETAGNE, pages 403-406, XP002009617 Y. TROTTIER ET AL.: "Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias." see the whole document</p> <p style="text-align: center;">---</p>	<p>1-6,13, 15-20, 23-25</p>
T	<p>HUMAN MOLECULAR GENETICS, vol. 5, no. 12, December 1996, OXFORD, GRANDE BRETAGNE, pages 1887-1892, XP002027564 G. STEVANIN ET AL.: "Screening for proteins with polyglutamine expansions in autosomal dominant cerebellar ataxias." see the whole document</p> <p style="text-align: center;">-----</p>	<p>16-20,24</p>

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/FR 96/01773

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 24  
because they relate to subject matter not required to be searched by this Authority, namely:  
Observation: although claim 24 (inasmuch as it refers to a method carried out in vivo) concerns a diagnostic method applied to the human or animal body, the search was carried out and based on the effects which have been attributed to the product/composition.
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 96/01773

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9501437 A	12-01-95	CA 2166117 A	12-01-95
		EP 0707647 A	24-04-96
		JP 9501049 T	04-02-97
-----			

Form PCT/ISA/210 (patent family annex) (July 1992)

## RAPPORT DE RECHERCHE INTERNATIONALE

Der e Internationale No

PCI/FR 96/01773

A. CLASSEMENT DE L'OBJET DE LA DEMANDE  
 CIB 6 C12N15/13 A61K39/395 A61K48/00 C07K16/18 C12N15/86  
 G01N33/577 G01N33/68 C12Q1/68 C12N15/12

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

## B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 6 C12N A61K C07K G01N C12Q

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

## C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	WO 95 01437 A (REGENTS OF THE UNIVERSITY OF MINNESOTA) 12 Janvier 1995 voir exemple IX voir revendications	21,22
X	THE EMBO JOURNAL, vol. 13, no. 5, 1 Mars 1994, OXFORD, GRANDE BRETAGNE, pages 1166-1175, XP002009616 A. LESCURE ET AL.: "The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters." cité dans la demande voir abrégé voir figure 1	23,24

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

## \* Catégories spéciales de documents cités:

- \*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- \*E\* document antérieur, mais publié à la date de dépôt international ou après cette date
- \*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- \*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- \*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- \*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- \*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- \*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- \*Z\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

14 Mars 1997

Date d'expédition du présent rapport de recherche internationale

24.03.97

Nom et adresse postale de l'administration chargée de la recherche internationale  
 Office Européen des Brevets, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tél. (+ 31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax (+ 31-70) 340-3016

Fonctionnaire autorisé

Nooij, F

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	<p>NATURE GENETICS, vol. 10, no. 1, Mai 1995, NEW YORK, NY, ATATS-UNIS, pages 104-110, XP000577139 Y. TROTTIER ET AL.: "Cellular localization of the Huntington's disease protein and discrimination of the normal and mutated form." voir le document en entier ---</p>	1,2,5, 13, 15-20, 23-25
A	<p>HUMAN MOLECULAR GENETICS, vol. 4, no. 3, Mars 1995, OXFORD, GRANDE BRETAGNE, pages 465-469, XP000577137 Y. JOU ET AL.: "Evidence from antibody studies that the CAG repeat in the Huntington disease gene is expressed in the protein." voir abrégé ---</p>	1,2,5, 13, 15-20, 23-25
A	<p>BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 209, no. 3, 26 Avril 1995, DULUTH, MN, ATATS-UNIS, pages 1119-1125, XP002009615 K. IDE ET AL.: "Abnormal gene product identified in Huntington's disease lymphocytes and brain." voir le document en entier ---</p>	1,2,5, 13, 15-20, 23-25
P,X	<p>NATURE, vol. 378, no. 6555, 23 Novembre 1995, LONDRES, GRANDE BRETAGNE, pages 403-406, XP002009617 Y. TROTTIER ET AL.: "Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias." voir le document en entier ---</p>	1-6,13, 15-20, 23-25
T	<p>HUMAN MOLECULAR GENETICS, vol. 5, no. 12, Décembre 1996, OXFORD, GRANDE BRETAGNE, pages 1887-1892, XP002027564 G. STEVANIN ET AL.: "Screening for proteins with polyglutamine expansions in autosomal dominant cerebellar ataxias." voir le document en entier -----</p>	16-20,24

# RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°

PCT/FR 96/01773

## Cadre I Observations - lorsqu'il a été estimé que certaines revendications ne pouvaient pas faire l'objet d'une recherche (suite du point 1 de la première feuille)

Conformément à l'article 17.2)a), certaines revendications n'ont pas fait l'objet d'une recherche pour les motifs suivants :

1. ☒ Les revendications n° 24  
se rapportent à un objet à l'égard duquel l'administration n'est pas tenue de procéder à la recherche, à savoir :  
  
Remarque: Bien que la revendication 24 (pour autant qu'il s'agit d'une méthode in vivo) concerne une méthode de diagnostic appliqué au corps humain/animal, la recherche a été effectuée et absée sur les effets imputés au produit/à la composition.
2. ☐ Les revendications n°  
se rapportent à des parties de la demande internationale qui ne remplissent pas suffisamment les conditions prescrites pour qu'une recherche significative puisse être effectuée, en particulier :
3. ☐ Les revendications n°  
sont des revendications dépendantes et ne sont pas rédigées conformément aux dispositions de la deuxième et de la troisième phrases de la règle 6.4.a).

## Cadre II Observations - lorsqu'il y a absence d'unité de l'invention (suite du point 2 de la première feuille)

L'administration chargée de la recherche internationale a trouvé plusieurs inventions dans la demande internationale, à savoir :

1. ☐ Comme toutes les taxes additionnelles ont été payées dans les délais par le déposant, le présent rapport de recherche internationale porte sur toutes les revendications pouvant faire l'objet d'une recherche.
2. ☐ Comme toutes les recherches portant sur les revendications qui s'y prêtaient ont pu être effectuées sans effort particulier justifiant une taxe additionnelle, l'administration n'a sollicité le paiement d'aucune taxe de cette nature.
3. ☐ Comme une partie seulement des taxes additionnelles demandées a été payée dans les délais par le déposant, le présent rapport de recherche internationale ne porte que sur les revendications pour lesquelles les taxes ont été payées, à savoir les revendications n° :
4. ☐ Aucune taxe additionnelle demandée n'a été payée dans les délais par le déposant. En conséquence, le présent rapport de recherche internationale ne porte que sur l'invention mentionnée en premier lieu dans les revendications; elle est couverte par les revendications n° :

Remarque quant à la réserve ☐ Les taxes additionnelles étaient accompagnées d'une réserve de la part du déposant.  
☐ Le paiement des taxes additionnelles n'était assorti d'aucune réserve.

## Renseignements relatifs aux : nombres de familles de brevets

PC1/FR 96/01773

Formulaire PCT/ISA/210 (annexe familles de brevets) (juillet 1992)

**THIS PAGE BLANK (USPTO)**